

Calibration

How to Weigh What You Know.
An Extension of the Book on Truth

Kai Froeb

About this translation

This translation was produced by machine using AI and has not yet been fully checked by a human. If you find any errors, please report them to us at feedback@reflexivitypress.com. Anyone who has bought the book and reports errors to us will receive the corrected version as an e-book free of charge.

Author
Kai Froeb
https://kai.froeb.net

Publisher
Reflexivity Press
São Martinho do Porto
Kai Froeb Edition

Latest version and notes on corrections:
https://reflexivitypress.com/en/books/froeb-kalibrierung
Feedback: feedback+en.kalibrierung.v1.0@reflexivitypress.com
Cover: https://reflexivitypress.com/en/about/cover-art/

Publication
2026-11-17
Translation: 2026-11-17
Version 1.0
Typeset with LaTeX (LuaHBTeX, Pandoc + Eisvogel)

Note on authorship
Written with the language model Claude (Anthropic). Conception, content and responsibility rest with the author. The translation was made with the help of AI language models. How the work was done is described in the essay “Philosophy in the Age of Machines That Write”: https://reflexivitypress.com/en/about/machines-that-write/

© Copyright 2026 by Reflexivity Press, São Martinho do Porto.
All rights reserved.

Zero Marginal Cost License 1.3 (ZMC 1.3)

Reader-friendly summary

What has no marginal cost should have no barriers.

You may

- **Read, download, and copy** the work free of charge.
- **Share the unaltered work digitally for noncommercial purposes**, with attribution to the author, source, and license.
- Convert the file format and make technically necessary accessibility adjustments without changing the content.
- **Print or commission individual copies and give them away** for personal, noncommercial use; direct costs may be reimbursed without markup or profit.
- **Translate the work** for your own personal, nonpublic, noncommercial use.
- Use the work individually for education and research, and quote it for noncommercial purposes in accordance with scholarly standards, without using a collection of excerpts as a substitute for the work.

Conditions

Use must remain **noncommercial**. When sharing the work publicly, identify the **author, source, and license**, wherever reasonably possible. You must not impose additional legal or technical restrictions or suggest endorsement by the author or publisher.

All other rights reserved

This license permits only the uses expressly stated. Any use beyond these permissions requires separate authorization from the rights holder, unless it is already permitted by law.

Statutory permissions and remuneration claims remain unaffected. Third-party material may be subject to separate terms.

Important: This summary is for guidance only and is not a standalone license. The full **ZMC 1.3** is the authoritative text:

https://reflexivitypress.com/en/zmc/



About the Author

Kai Froeb (born 1963) is a German philosopher and IT expert living in São Martinho do Porto, Portugal. His work combines more than 35 years of systematic Hegel research with the possibilities of artificial intelligence, developing a new form of philosophical collaboration.

His thinking is radically interdisciplinary. He studied mathematics and computer science at the University of Erlangen–Nuremberg, business administration at EBS, and cultural studies and law at the University of Hagen. At the same time, he pursued intensive general studies in philosophy, sociology and history. He deliberately conducted his studies without orienting them towards academic degrees, preserving the freedom to follow an independent path of thought and research beyond the boundaries of university disciplines.

Since the late 1990s, Froeb has edited and maintained **hegel.net**. Available in twelve languages, the website ranked among the leading results in international Google searches for Hegel for many years and developed into one of the central online resources on Hegel's philosophy. It is probably still the most extensive multilingual website devoted to Hegel on the internet.

In connection with **hegel.net** and **hegel-system.de**, Froeb also created an extensive, publicly accessible archive of primary sources. The German-language collection alone offers 506 digitised original editions of works by Hegel and nineteenth-century Hegelians for free download. Further collections in other languages include 19 original works in English.

More than 40 years of programming experience, beginning in 1980, form another methodological foundation of his work. His first program was a reimplementation of Joseph Weizenbaum's ELIZA, thus completing a circle from the earliest experiments in artificial intelligence to his productive collaboration with Claude today. Froeb uses his IT expertise not merely as a technical resource but also as an instrument of thought: he analyses social and philosophical structures with the precision of a software architect.

Froeb is developing an independent philosophical body of work that stands in the tradition of Hegel while attempting to continue that tradition under contemporary conditions. At its centre is his own performative approach to truth, which forms the common foundation of his philosophical books. Truth appears here not merely as the object of propositions, but as a movement that must also prove itself in the form of thinking and presentation.

His aim is demystification through conceptual clarification: to penetrate every boundary through thought, including the boundary of what initially appears inaccessible. In his work, form follows content—structure, concepts and method are determined by the subject matter itself. At the same time, he seeks accessibility: to present demanding philosophy as clearly as possible, in the spirit of the best Anglo-American scholarly tradition, without simplifying it.

His guiding motif is Hegel's formula: "The truth is the whole"—the search for a form of integration that need not repress or exclude anything.

CONTENTS

Preface 1

Part I: What Is Missing 3
Chapter 1: The Map and the Territory . . . 3
1.1 The Square in Baghdad . . . 3
1.2 What Every Representation Leaves Out . . . 4
1.3 The Elephant with a Part Missing . . . 4
1.4 What Calibrating Means . . . 5
Summary . . . 6
Chapter 2: What Is Missing and the Wrong Weight . . . 6
2.1 A Gap Announces Itself When One Needs It . . . 6
2.2 A Weight Lies in No Single Part . . . 7
2.3 Where One Knows Little, Chance Carries Weight . . . 8
2.4 Two Checks . . . 9
Summary . . . 9

Part II: Numbers 11
Chapter 3: The Base Rate . . . 11
3.1 A Test . . . 11
3.2 The Description Crowds Out the Number . . . 12
3.3 Which Group? . . . 12
3.4 Where One Gets a Base Rate . . . 13
Summary . . . 13
Chapter 4: Denominator and Scale . . . 14
4.1 Twice as High . . . 14
4.2 Added Up . . . 15
4.3 Where the Threshold Lies . . . 15
4.4 Who Is in the Denominator . . . 16
Summary . . . 17
Chapter 5: What Gets Counted . . . 18
5.1 Two Million Answers . . . 18
5.2 What Gets Published . . . 18
5.3 Two Ways of Counting in War . . . 19
5.4 The Question for Every Number . . . 20
Summary . . . 21

Part III: Foreign Countries **23**
Chapter 6: What One Knows from Afar . . . 23
6.1 Every Channel Has Its Filter . . . 23
6.2 Accounts Are Filtered . . . 23
6.3 What Cannot Be Shown . . . 24
6.4 What Once Held . . . 24
6.5 Calibrating from Afar . . . 25
Summary . . . 25
Chapter 7: On the Spot and the Regions . . . 26
7.1 What Comes Across Only on the Spot . . . 26
7.2 Country, Region, City, Neighborhood . . . 26
7.3 Calibrating a Stay . . . 27
Summary . . . 28

Part IV: Weight as a Means **29**
Chapter 8: Propaganda . . . 29
8.1 The Lie and Its Price . . . 29
8.2 The Rule: Selection . . . 30
8.3 Balance as Staging . . . 30
8.4 What One Thinks About . . . 31
Summary . . . 32
Chapter 9: Flooding . . . 32
9.1 Quantity Instead of Persuasion . . . 32
9.2 What Quantity Does to the Measure . . . 33
9.3 What Keeps the Measure Stable . . . 34
Summary . . . 34
Chapter 10: Cognitive Warfare . . . 35
10.1 The Security Argument . . . 35
10.2 Two Ways . . . 36
10.3 Why Steering Defends Less Well . . . 36
10.4 Why Calibration Defends Better . . . 37
10.5 The Limit in an Emergency . . . 38
Summary . . . 39

Part V: Weighting Together **41**
Chapter 11: Committees, Experts, Consensus . . . 41
11.1 Eight Hundred Estimates . . . 41
11.2 The Knowledge in the Room . . . 42
11.3 The Appointed Dissent . . . 42
11.4 Consensus About What . . . 43
Summary . . . 44
Chapter 12: Artificial Intelligence . . . 44
12.1 What Was Written . . . 44
12.2 The Tendency to Agree . . . 45

12.3 One Channel for Many 45
12.4 What It Is Good For 46
Summary 47

Part VI: Calibrating Oneself **49**
Chapter 13: One's Own Scale 49
13.1 When a Judgment Is Calibrated 49
13.2 A Test 49
13.3 Keeping a Record 50
13.4 The Symmetry Test 51
13.5 The Strongest Opposing Side 51
Summary 52
Chapter 14: Calibrating Together 52
14.1 Why Several Voices Are Needed 52
14.2 Weighting Disputed Questions Oneself 53
14.3 No False Middle 54
14.4 Weighting with Others 54
Summary 55
Chapter 15: Inverted Calibration 55
15.1 Knowledge That Works in Both Directions 55
15.2 The Questions of This Book, Inverted 56
15.3 How One Recognizes the Inversion 57
15.4 How to Protect Oneself and Others 57
Summary 58
Chapter 16: Limits 58
16.1 Where No Feedback Comes 58
16.2 What Is Missing and Does Not Show Itself 59
16.3 No Main Switch, No Conclusion 59
16.4 This Book as a Selection 60
16.5 The Step Back 60
Summary 60

PREFACE

Whoever reads a newspaper, watches a documentary, or listens to a friend back from a journey mostly hears what is true. Even so, the picture he gains from it can be skewed. Something is missing, or something weighs too much, and none of the sentences from which the picture arose was false. This book is about how to deal with that: how to check the weight of the parts from which one forms a picture, and how to look for what is missing and cannot be seen.

The book extends an earlier one about truth, which treats this question in one chapter. There the subject was that we check, how we check, and why checking makes free. Calibration was the hardest case in it: checking what is not there at all. Here it is worked out, on numbers, on foreign countries, on propaganda, on institutions, and on one's own judgment. One does not need the earlier book for this. What this one takes over from it is stated in one sentence at the place where it is needed; where the other book develops a thought further, a line at the end of the section says so.

The examples come from many countries, and none of them is needed to understand the thought; they show it in what is real and should therefore be correct. Their sources are in the footnotes. Where a matter is disputed, the strongest version of each side is given, together with the way in which one can weigh it oneself. At the end of each chapter there is a summary. Whoever reads only the summaries knows the course of the book; whoever doubts a point finds the reasons before it.

Further reading: *Truth* (2025)

PART I: WHAT IS MISSING

Chapter 1: The Map and the Territory

1.1 The Square in Baghdad

On April 9, 2003, three weeks after the start of the Iraq War, a large statue of Saddam Hussein was pulled from its pedestal in Firdos Square in Baghdad. Television stations all over the world broadcast it, many for hours and again and again. The pictures showed cheering men around the pedestal, and whoever saw them understood: Baghdad is celebrating its liberation. Most of the pictures were close-ups. The rare wide shots showed a square that was largely empty. A reporter who later went through the footage and the accounts of those involved estimated the crowd at a few hundred people at most, a quarter to a half of whom were journalists and soldiers. The statue had been pulled down by an armored recovery vehicle of the United States Marine Corps.[1]

Nothing in the close-ups was faked. The statue fell, the men cheered, and some of them may have suffered under the man who was toppled. What became false was only what the viewer made of it: that the city had gathered in this square. The camera had prepared this step without asserting anything. It showed a frame, and because the picture was full to the edges, it seemed to have no edge. The farther one moved away from the pedestal, the more one saw, and the less impressive was what one saw.

Every camera has a frame, even the best, and whoever chooses the frame decides not about the truth of the picture but about its weight. Whether it is a square or a country, a disease or a war: what one sees is a part, and the part does not say of itself how large the whole is from which it comes.

1 Peter Maass, "The Toppling. How the Media Inflated a Minor Moment in a Long War," *ProPublica*, January 3, 2011, published jointly with *The New Yorker*: "Very few Iraqis were there"; the crowd "does not seem to have exceeded several hundred at its largest," a quarter to a half of them journalists or Marines; close-ups created "an illusion of wall-to-wall enthusiasm," while in the rare wide shots the square was "mostly empty."

1.2 What Every Representation Leaves Out

One might think the evil is the frame, and an honest representation would simply show everything. That is impossible; it would not even be desirable. A map that left out nothing would have to be as large as the country; it would cover it and be of use to no one. A road map leaves out the contour lines, a hiking map the gas stations, a weather map almost everything except pressure, clouds, and wind, and precisely for that reason each serves its purpose. Condensing is what it achieves. Errors arise from it only when what was left out would have been exactly what mattered for the question one puts to the map.

The same holds for every measurement. A thermometer measures the air at one spot, not in the valley and not on the mountain. A statistic counts what was recorded: the reported cases, not the ones that occurred; the answers of those who answered, not the opinion of everyone. And the very question with which one investigates something already fixes what one can find. Whoever asks how many accidents a road has learns nothing about the accidents that happen on the bypass because drivers avoid the dangerous road.

What has been left out, then, does not lie in a place where one could go and look. It lies outside the frame of the picture, outside the measurement, outside the question, outside what has been established at all so far. One does not see it, and so one does not see that it is missing either. A blank spot on the map stands out. What was never drawn in does not.

Lewis Carroll put the thought into a joke in 1893: a scholar reports that in his country they made a map on the scale of a mile to the mile but never spread it out, because the farmers feared it would cover the country and shut out the sunlight; now they use the country itself as its own map, and it does nearly as well.[2] Jorge Luis Borges later turned this into a short piece of prose in which the cartographers of an empire draw a map as large as the empire.[3]

1.3 The Elephant with a Part Missing

An old story tells of six blind men who touch an elephant. One grasps the leg and says an elephant is like a pillar, the second the trunk and says it is like a snake,

2 Lewis Carroll, *Sylvie and Bruno Concluded*, London 1893, ch. 11 (“The Man in the Moon”): “We actually made a map of the country, on the scale of a mile to the mile!” – “the farmers objected: they said it would cover the whole country, and shut out the sunlight!” – “So we now use the country itself, as its own map, and I assure you it does nearly as well.”

3 Jorge Luis Borges, “Del rigor en la ciencia”; on the place of this text, Ivan Almeida, “Borges à la carte (tres citas de Baudrillard),” *Variaciones Borges* 25 (2008).

the third the ear and says it is like a fan, and so on. They quarrel until it turns out that each has felt something true and only all the reports together make up the animal. The story is usually told as a lesson about different perspectives that complement one another.

The problem of this book is the same, with one difference: here something is missing. Suppose only three of the six come back: those with the tusk, the leg, and the trunk. Their reports are correct, they do not contradict one another, and whoever puts them together gets an animal made of spear, pillar, and snake that is more threatening than the elephant. The ear, the belly, and the tail are missing, and no one misses them, because no one knows that they exist.

In the story each knows that he has touched only a part; that is why he can come to an understanding with the others. In everyday life the part usually presents itself as the whole. The close-up from Firdos Square did not say that it was a close-up, and the documentary about one neighborhood does not say that the other neighborhoods are different. That makes completing the picture harder than the story shows. One first has to realize that there is something to complete.

Further reading: *Truth* (2025), Ch. 5 (the parable of the elephant), *Thinking – Ordering Without Losing* (2025), Ch. 2 and 14 (condensing and what is lost in it)

1.4 What Calibrating Means

In measurement technology, to calibrate means to compare the display of an instrument with a known measure: one puts a weight one knows on the scale and looks at how far it deviates. A thermometer is checked at two states whose temperature is fixed, melting ice and boiling water, and one reads off by how much it is off at both points. Two things belong to this. One needs a measure that one trusts more than the instrument. And the instrument cannot check itself against its own readings; a scale that measures only what it has itself weighed never finds its deviation.

In judging it is the same. Everyone brings along a measure of what is frequent and what is rare, what is ordinary and what is extraordinary, what is important and what is secondary, what most people think and what only a few think. This measure is not a proposition one defends but a feel for magnitudes, and it has settled out of everything one has seen, read, and heard. To calibrate here means to check this measure against something that does not come from the same stream: against a count, against a second, independent path, against one's own

observation, against a case whose outcome one learns later. Where it deviates, one readjusts it.

Checking individual statements does not settle this. Whoever checked every report about Firdos Square one by one would find them all confirmed. What would need checking is the size he gives them together. That is why calibration is also directed at what is not there, and that is why it is harder than checking what is said.

Whoever asks what he is missing, however, already has the easier part of the task before him. Harder is what one does not notice as a gap at all: the wrong weight of what one has right before one's eyes.

Further reading: *Truth*, Ch. 19 (the core of this book in one chapter)

Summary

Every representation leaves something out, and that is not a weakness but the condition of its usefulness: a map that contained everything would be as large as the country. Camera, map, measurement, and question show a frame, and the frame does not say of itself how large the whole is. In Firdos Square in 2003 the close-ups showed what was true and let the world believe that a whole city was celebrating where a few hundred people at most were standing. The problem resembles that of the blind men who touch an elephant, except that here a part is missing and no one knows that it exists; in everyday life, moreover, the part presents itself as the whole. To calibrate means, as in measurement technology, to check one's own measure of frequency, size, and weight against something that does not come from the same stream, and to readjust it where it deviates. That is why it is also directed at what is not there.

Chapter 2: What Is Missing and the Wrong Weight

2.1 A Gap Announces Itself When One Needs It

Whoever is asked what people in Mongolia live on, what music people listen to in Ghana, or how elections work in Peru often notices at that moment that he knows nothing about it. The gap was there before, but it did not announce itself. Now it does, because a question meets it and finds no answer. What to do then is not hard: read up, ask someone who has been there, look for a number. A gap one has noticed is a task, and for most tasks of this kind there are ways.

Harder to notice is a gap in what one is investigating. In the Second World War the American military asked statisticians how vulnerable its bombers were at which points. The data came from the aircraft that returned, with bullet holes in the fuselage and wings. But precisely these aircraft show where hits can be survived. The others, which had been hit at the sensitive points, had not returned and did not appear in the data. Whoever counts only what is in front of him reinforces the places where nothing bad happens even now. The way out was found by whoever asked where the data came from and what it meant that part of the aircraft was missing from them.

The story is usually attributed to the statistician Abraham Wald and embellished in its best-known form. What is documented is that in 1943 Wald calculated, in a series of memoranda, how one can infer the vulnerability of all aircraft from the damage to those that returned. The anecdote that he had the armor placed where the returning planes had no holes is not in these memoranda.[4] This changes nothing about the thought: the gap became visible only when someone asked about what was missing in the material, not about what was in it.

2.2 A Weight Lies in No Single Part

Whoever has seen five documentaries about a country, one about violence in a big city, one about a strike, one about a drought, one about a festival, and one about a technology company, can check each one. Whether the violence happened there, whether the strike took place, can be established. What cannot be checked in any of the five is their relation to one another and to everything that was not reported on: whether violence is frequent in this country or rare, whether the strike is the rule or the exception. A weight is a relation. It lies in no single part, and that is why one does not find it by checking the parts one by one.

Hence the proposition on which this book rests: that something is missing entirely is easier to find and to remedy than that what is shown is wrongly weighted. The gap announces itself when a question finds no answer. The wrong weight does not announce itself, for every question about a single piece finds its answer – one that is correct. One would have to have the whole before one to see that the

4 Abraham Wald, memoranda for the Statistical Research Group of Columbia University (1943), published by the Center for Naval Analyses in 1980; Marc Mangel, Francisco J. Samaniego, “Abraham Wald’s Work on Aircraft Survivability,” *Journal of the American Statistical Association* 79 (1984). On distinguishing evidence from legend, Bill Casselman, “The Legend of Abraham Wald,” Feature Column of the American Mathematical Society (2016): the memoranda contain no recommendation on armor, “there is very little evidence for the best bits.”

pieces in it have turned out too large or too small, and one does not have the whole; otherwise one would not need the pieces.

Where, then, does one get a weight at all? Mostly from the frequency with which one has encountered something. What one has often heard one takes to be frequent; what comes easily to mind, to be widespread. For much in everyday life that is a good rule of thumb, for what occurs frequently is also encountered frequently. It fails as soon as there is a filter between the world and one's own memory that selects by something other than frequency. The news selects by what is new and unusual; a plane crash makes the news because it is rare. Whoever draws his sense of frequency from the news is therefore measuring with a measure that enlarges the rare. The psychologists Amos Tversky and Daniel Kahneman called this tendency to estimate frequency by how easily examples come to mind availability.[5]

2.3 Where One Knows Little, Chance Carries Weight

Where one's own experience is thin, the measure fills up with whatever happens to be there. In an experiment described by the same researchers, the experimenters spun a wheel of fortune before the participants' eyes that yielded a number between 0 and 100. Then they asked whether the share of African states among the members of the United Nations was higher or lower than this number, and after that, how high it was. Those who had seen 10 estimated 25 percent on average, those who had seen 65, 45 percent.[6] The number from the wheel had nothing to do with Africa. Everyone knew that. It carried weight all the same, because little else was there that could have carried weight.

That the experiment asks about Africa of all places suits it. Of a whole continent with more than fifty states, many people outside it have only a rough picture, and a rough picture has no weight of its own that could hold out against a random number. What holds for the share in the United Nations holds just as much for the poverty, the violence, the life expectancy, the economic growth of a distant country. The first number one encounters, the most striking story, the last picture become the measure until something better takes their place.

---

5 Amos Tversky, Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science* 185 (1974), pp. 1124–1131, here p. 1127: "one may assess the risk of heart attack among middle-aged people by recalling such occurrences among one's acquaintances … This judgmental heuristic is called availability."

6 Tversky, Kahneman, "Judgment under Uncertainty," *Science* 185 (1974), p. 1128: "the median estimates of the percentage of African countries in the United Nations were 25 and 45 for groups that received 10 and 65, respectively, as starting points."

Experiment 1: Where Your Weights Come From

Think of a country in which you have never been and about which you nonetheless think you know something. Write down three things that, by your impression, are frequent or typical there. Then ask about each: Where did I get this – from a film, from the news, from someone who was there? How often have I encountered it, and by how many different paths? And would I have heard of it if it were rare there?

Probably the three come from a few sources, and for some of them the last question becomes uncomfortable. A film shows the unusual because it carries a story, and a news report reports what departs from the usual. What you take to be typical may be typical. But your reason for it is often only that you have encountered it often, and that was decided by a filter that asks about something other than how frequent something is.

2.4 Two Checks

That is why there are two checks that ask different questions. The first asks of each statement whether it is true: who makes it, how he knows it, whether independent sources say the same. Against the lie it is the right tool, long tried and tested. The second asks about the distribution: whether what is shown is representative, what it does not show, and why precisely this lies before one. It needs something the first does not need, namely a measure of the whole against which the parts can be measured. Where one gets such a measure when one does not have the whole before one can be asked most clearly about numbers.

Further reading: *Truth* (2025), Ch. 19 (the first and the second level)

Summary

That something is missing entirely announces itself as soon as a question finds no answer; then one can read up, ask, count. Harder to find is a gap in the material itself, as with the bombers of the Second World War, whose hits were seen only on the aircraft that returned. Hardest of all is the wrong weight, for a weight is a relation: it lies in no single piece, and whoever checks each piece finds each confirmed. Weights mostly arise from the frequency with which one encounters something. That holds as long as there is no filter between world and memory that selects by something else; the news selects the unusual. Where one knows little,

chance fills the measure, like the number from the wheel of fortune that shifted the estimate about Africa. That is why, besides checking the individual statement, a second check is needed that asks about the distribution, and it needs a measure of the whole.

PART II: NUMBERS

Chapter 3: The Base Rate

3.1 A Test

Experiment 2: A Positive Result

A disease occurs in a population in one person in a thousand. A test detects it in 99 out of 100 sick people. In healthy people it falsely gives a positive result in 5 out of 100 cases. Someone from this population is tested without any particular reason, and the test is positive. How probable is it that he has the disease? Make an estimate before you read on.
Then work it out for a hundred thousand people. A hundred of them are sick, and the test is positive for 99 of them. 99,900 are healthy, and for five out of a hundred of these, that is, for 4,995, it is positive as well. That makes 5,094 people who test positive, and 99 of them are sick. That is just under two percent.

Whoever estimated ninety percent or more is in large company; in tasks of this kind many people estimate far too high as long as the data are given as percentages. Whoever receives the same data as frequencies, as "a hundred out of a hundred thousand," more often arrives at the right magnitude.[1] The calculation is the same in both versions. What differs is whether it leaves the base rate visible.

The base rate here is the share of sick people in the group from which the person tested comes, one in a thousand. It decides what the result is worth. The 99 percent of the test force themselves on one because they belong to the result; the one in a thousand stands beside it like a side issue. Yet it weighs more. With a rare disease, even a few false alarms among many healthy people add up to

1 Gerd Gigerenzer, Ulrich Hoffrage, "How to Improve Bayesian Reasoning without Instruction: Frequency Formats," *Psychological Review* 102 (1995), pp. 684–704.

far more than the few correct hits among the sick. The weight of the result thus depends on a number that does not appear in the result itself.

3.2 The Description Crowds Out the Number

In another experiment, participants were given short descriptions of people who had supposedly been drawn from a group of engineers and lawyers. Some were told that the group consisted of 70 engineers and 30 lawyers, the others the reverse. They were to estimate how probable it was that the person described was an engineer. Without a description, the participants correctly started from the composition of the group. With a description, the composition hardly appeared in their estimates at all.[2]

A description is vivid; a base rate is a number without a face. In memory the description has the advantage, and so it crowds out the number, even where the number says more. This is the same movement as with the documentary and the country from the previous chapter, only more precisely measurable: the single case, vividly told, takes for itself the weight that would belong to the distribution.

3.3 Which Group?

A base rate is always the rate in a group, and which group is meant is not fixed. If someone comes to the doctor with complaints that fit the disease, he no longer comes from the whole population but from the group of those who have such complaints. In it the disease is more frequent, perhaps one in twenty. The same test with the same positive result then means something quite different: if one calculates as above, about every second person who tests positive is now really sick. Nothing has changed in the test, only in the group to which one applies it.

The choice of group is thus itself a weighting; there is no rule that decides it for every question. One takes the narrowest group that resembles the case and for which there is a reliable count. If it is too wide, what distinguishes the case becomes blurred; if it is too narrow, there is no number any more, or it rests on three cases. Whoever reads about a crime and wants to ask how frequent something like this is must decide whether he means the rate in a country, in a city, in a neighborhood, or among people of a certain age. Each of these numbers says something different. It becomes contentious where the choice of group

2 Tversky, Kahneman, "Judgment under Uncertainty," *Science* 185 (1974), p. 1125: the group "consisted of 70 engineers and 30 lawyers" (or the reverse); "prior probabilities were effectively ignored when a description was introduced."

suggests a judgment one already had beforehand. Then it pays to set all the obvious groups side by side and look at how far apart they lie.

3.4 Where One Gets a Base Rate

In everyday life one rarely has a number at hand, and the sense of frequency one has instead is exactly what would need checking. That is why one gains a base rate not from impressions but from counts: from official statistics, from surveys that state whom they asked, from one's own records of cases whose outcome one knows. Even a rough count is better than none. Whoever knows only whether something occurs in one in ten or one in ten thousand already knows whether a result weighs much or little.

Precisely with reports of something unusual, the base rate must be asked about twice. How frequent is the event itself, and how frequently are events of this kind reported? A report proves that it happened. But its frequency in the news says not how often it happens but how often it is considered worth reporting, and that depends on its rarity, on its horror, and on the question of whom it serves at the moment.

A low base rate refutes nothing. Rare things happen, and a positive test is a reason to check further even with a rare disease. The base rate determines how much a single result shifts and how much more one needs before one may be sure. In the eighteenth century Thomas Bayes put the thought into a formula that states how a new result changes a prior probability. One does not need its mathematics in everyday life. It is enough not to leave out the base rate.

The base rate already shows that a number says something only when one knows what it refers to. The same holds for every fraction, every share, and every comparison, and there the denominator decides.

Further reading: *Truth* (2025), Ch. 19 (economy of checking: base rate, stakes, track record of the source)

Summary

A test that detects ninety-nine out of a hundred sick people and gives a positive result for five out of a hundred healthy people yields, for a disease that affects one in a thousand, only about two percent sick people among those who test positive. The weight of a result depends on the base rate, which does not appear in the result itself. Calculated in frequencies instead of percentages, it stays visible. A vivid description easily crowds out the base rate, even where it says little. Every

base rate is the rate in a group, and the choice of group is itself a weighting: for a patient with complaints, the same test means something quite different. One gains base rates from counts, not from impressions; with reports, one must ask how frequent the event is and how frequently it is reported. A low base rate refutes nothing; it determines how much a single result shifts.

Chapter 4: Denominator and Scale

4.1 Twice as High

On October 19, 1995, the British medicines regulator warned doctors and the public that certain newer contraceptive pills doubled the risk of venous thrombosis compared with older ones. Many women stopped taking the pill. In the months that followed, the number of pregnancies in England and Wales rose by about seven percent, that of abortions by six to nine percent.[3]

The warning was not invented. The studies on which it rested had found a difference, and an authority that knows of it and keeps silent fails in its duty. What determined the effect was the form. "Twice as high" is a ratio of two numbers that says nothing about how large these numbers are. A later, more precise analysis arrived at about 20 cases per hundred thousand women per year with the older preparations and about 40 with the newer ones.[4] Twice as high, and in absolute numbers a difference of two cases among ten thousand women. Whether that is a lot can be said only by comparison, for instance with the consequences of an unwanted pregnancy, and the warning did not include this comparison.

The relative risk is the weight without its object. It is suitable for comparing two possibilities; it deceives where someone wants to know how much is at stake. Whoever reads such a report therefore asks about the denominator: twice as high as what, among how many? Whoever writes it owes both, the ratio and the numbers from which it is formed. The same holds in the opposite direction. A

3 Emilia Del Bono, Marco Francesconi, Nicky Best, *Health Information and Health Outcomes: An Application of the Regression Discontinuity Design to the 1995 UK Contraceptive Pill Scare Case*, Discussion Paper 696, Department of Economics, University of Essex 2011: warning of the Committee on Safety of Medicines of October 19, 1995 ("twice the risk of venous thromboembolism"); increase in conceptions of 5.7 to 7.4 percent and in abortions of 6.4 to 9.5 percent in the following months.

4 D. Hollander, "Most Refined Analysis to Date Confirms Link Between Third-Generation Pills and Venous Thromboembolism," *Perspectives on Sexual and Reproductive Health* 33 (2001), no. 3, pp. 133–134: crude rates of "37–41 cases per 100,000 person-years" for third-generation pills and "20–23 per 100,000" for pills containing levonorgestrel.

halving that leads from two cases to one in ten thousand is little for the individual; for a country of fifty million people it is five thousand.

4.2 Added Up

In 1973, 8,442 men and 4,321 women applied for graduate admission at the University of California, Berkeley. About 44 percent of the men and 35 percent of the women were admitted. That looked like discrimination, and the university had it investigated. Department by department, that is, where the decisions were actually made, no general discrimination against women was found; if there was a bias, it was rather to the disadvantage of the men. The difference in the whole came from the fact that women had more often applied to departments that admitted only a few applicants, men more often to departments that admitted many.[5]

Both numbers were correct, the one for the whole and those for the parts, and they said different things. Whoever adds up weights the parts by their size, and if the groups are distributed differently across the parts, the whole can point in the opposite direction from every part. Which number is the right one depends on the question. Whoever wants to know whether the departments discriminated in their selection must look at the departments, for that is where the decisions were made. Whoever wants to know why women applied to the more selective departments and whether another disadvantage lies behind it, earlier at school or later at work, is asking a new question, and for it the number for the whole is the beginning, not the end.

Added-up numbers are everywhere: the mortality of a hospital that admits more serious cases than another, the average wage in a country whose population is aging, the success rate of an agency that handles the easy cases first. In each of these cases one must ask what parts the whole consists of and whether their shares have shifted.

4.3 Where the Threshold Lies

According to the World Bank's figures, in 1990 about 38 percent of people lived on less than 2.15 dollars a day, converted to the purchasing power of a dollar in the United States. In 2019 it was 8.5 percent. According to the same figures, in

5 P. J. Bickel, E. A. Hammel, J. W. O'Connell, "Sex Bias in Graduate Admissions: Data from Berkeley," *Science* 187 (1975), pp. 398–404. Figures (8,442 men, 4,321 women; about 44 and 35 percent) as given in David Freedman, Robert Pisani, Roger Purves, *Statistics*, 3rd ed., New York 1998, ch. 2.

2024 44 percent of humanity still lived on less than 6.85 dollars a day, and the number of people below this threshold has hardly changed since 1990, because the growth of the population offset the decline in their share.[6]

Whoever says that the world's poverty has fallen dramatically within a generation is right. Whoever says that almost half of humanity still lives in poverty and that there are as many as in 1990 is right too. Both take the same statistic and choose a different threshold, and the choice of threshold decides which picture emerges. Neither of the two thresholds is wrong. The low one counts those who do not have enough to survive; the higher one, those who have hardly more than that. How one weights them depends on what one wants to know: whether the misery of hunger is receding, or whether the world has become a place where most people can live well.

Whoever reads a number that refers to a threshold – a poverty line, a limit value, a definition of success or of violence – therefore asks where the threshold lies, who set it, and what shows when one moves it. If the picture stays the same at different thresholds, it is solid. If it reverses, the number says less about the world than about the choice of threshold.

4.4 Who Is in the Denominator

In 2013 researchers analyzed almost twelve thousand abstracts of scientific papers on global warming that had appeared between 1991 and 2011. 66.4 percent expressed no position on whether humans cause the warming. 32.6 percent affirmed it, 0.7 percent denied it, 0.3 percent considered the cause uncertain. Of those that took any position on the cause at all, 97.1 percent affirmed that humans cause the warming.[7]

Two sentences can be formed from the same study. "97 percent of experts agree" (DE: „97 Prozent der Fachleute sind sich einig") takes as its denominator only the papers that take a position. "Only a third of the papers affirm the human cause" (DE: „Nur ein Drittel der Arbeiten bejaht die menschliche Ursache") takes all of them. The first reading, in its strongest form, says: whoever writes in a

6 Daniel Gerszon Mahler, "Progress, Challenges in Ending Extreme Poverty," World Bank Data Blog, July 12, 2023 (decline in the share below 2.15 dollars "from about 38 percent to 8.5 percent" between 1990 and 2019); World Bank, press release on the report *Poverty, Prosperity, and Planet 2024*, October 15, 2024 (in 2024 almost 700 million people or 8.5 percent below 2.15 dollars, 44 percent below 6.85 dollars; number below 6.85 dollars hardly changed since 1990).

7 John Cook et al., "Quantifying the Consensus on Anthropogenic Global Warming in the Scientific Literature," *Environmental Research Letters* 8 (2013), 024024, doi:10.1088/1748-9326/8/2/024024 (11,944 abstracts; 66.4, 32.6, 0.7, and 0.3 percent; 97.1 percent of the abstracts with a position).

scientific journal about the consequences of warming for glaciers does not repeat its cause in every abstract, any more than a geologist mentions in every paper that the continents drift. The critics, in their strongest form, reply that one may not infer agreement from silence, and that the sentence with the 97 percent is understood as a count of people while it counts texts.

Both objections hit something. The remedy is the same in both cases: write down the denominator. “Among the papers that express a view on the cause” (DE: „Unter den Arbeiten, die sich zur Ursache äußern“) is a precise statement, and one can check it. What it does not say is what exactly there is agreement about, how strongly humans are warming the earth, and what follows from it. That is a question of its own; it belongs to the bodies that establish a consensus.

A denominator is an answer to the question of which whole a part is meant to be part of. It becomes more difficult when what is counted has already been selected before anyone calculates.

Further reading: *Thinking – Ordering Without Losing* (2025), Ch. 10 (a number read as the wrong kind of connection)

Summary

A number carries weight only when one knows what it refers to. The British pill warning of 1995 correctly said “twice as high” and concealed nothing, but it gave the ratio without its magnitudes: about twenty against forty cases per hundred thousand women per year. Afterward, pregnancies and abortions rose. Whoever adds up weights the parts by their size; at Berkeley in 1973 the whole showed a disadvantage for women that did not occur in the departments where the decisions were made, and which number counts depends on the question. The choice of a threshold decides whether the world’s poverty has fallen sharply or hardly at all: both are in the World Bank’s figures. And from the same analysis of scientific papers on the climate follow “97 percent” and “a third,” depending on who is in the denominator. The remedy each time is to write down the denominator, the magnitudes, and the threshold, and to see whether the picture holds when one changes them.

Chapter 5: What Gets Counted

5.1 Two Million Answers

Before the 1936 presidential election in the United States, the magazine *Literary Digest* sent ballots to ten million people. 2.4 million sent them back filled in. According to this enormous poll, the challenger Alf Landon would beat President Franklin D. Roosevelt. Roosevelt won by a wide margin. The magazine had taken the addresses mainly from telephone directories and lists of car owners, and only just under one in four had answered.[8]

Two million answers are more than any poll today collects. The quantity did not make the error smaller; it only made it more precise. Those on these lists were on average better off than the rest, and those who took the trouble to answer differed once more from those who did not. Two filters lay between the electorate and the result, and neither of them could be offset by more ballots. A small sample, drawn at random from everyone, would have hit the result better.

From this follows a question one can put to every number about opinions, experiences, or conditions: who had the chance to be counted at all, and who used it? An online survey reaches those who are online; the comments under an article come from those who comment; the ratings of a hotel from those who were especially satisfied or especially annoyed. Each of these numbers can be correct and yet hold only for those counted.

5.2 What Gets Published

Before a drug is approved in the United States, the manufacturer must submit to the regulatory authority all the studies it has registered for this purpose, including the unsuccessful ones. In 2008 researchers compared, for a number of antidepressants, what lay with the authority and what appeared in scientific journals. The authority had 74 studies. By its assessment 38 were positive, and 37 of these were published. Of the remaining 36, which had turned out negative or questionable, 22 did not appear and 11 appeared in a way that made them

8 Ten million ballots sent, 2.4 million returned, addresses "predominantly from telephone directories and automobile registration records," response rate 24 percent: Matthew J. Salganik, *Bit by Bit. Social Research in the Digital Age*, Princeton 2018, ch. 3.3.1 ("Representation"). In detail, Peverill Squire, "Why the 1936 Literary Digest Poll Failed," *Public Opinion Quarterly* 52 (1988), pp. 125–133.

look positive. In the published literature 94 percent of the studies looked positive; by the authority's assessment it was 51 percent.[9]

No single paper had to be false for this. Whoever read the literature read studies that had taken place, with results that had been measured. The distortion lay in the selection of what appeared, and it hit precisely the careful reader who reads many studies and summarizes their results. He got a picture according to which the drugs worked considerably better than the complete data showed.

This could be discovered only because there was a second register that had been set up before the result: the submissions to the authority. Therein lies a general means of calibration. A stock that was filtered by the result can be checked against a stock that was fixed beforehand. For the same reason the major medical journals have required since 2005 that clinical trials be publicly registered before they begin; otherwise they do not print the results.[10]

The thought reaches beyond medicine. Whoever wants to know whether he has the whole before him or a selection looks for such a list: for the plan of an investigation before its result, for the applications alongside the approvals, for all cases alongside those reported.

5.3 Two Ways of Counting in War

How many people a war kills is one of the most important numbers there is about it, and one of the hardest to determine. It can be seen in the Iraq War after 2003, because it is over and has been thoroughly investigated. An independent group counted the dead reported by the media and the authorities, case by case. A survey of 9,345 households, presented by the World Health Organization in 2008, estimated the number of those violently killed from March 2003 to June 2006 at 151,000, with a range of 104,000 to 223,000. That was about three times as many as the count from reports. A smaller household survey from 2006, with 1,849 households, had arrived at around 600,000 violent deaths for a similar period, about four times as many as the larger one.[11]

---

9 Erick H. Turner et al., "Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy," *New England Journal of Medicine* 358 (2008), pp. 252–260; the figures follow the reanalysis by Santiago G. Moreno et al., "Novel Methods to Deal with Publication Biases," *BMJ* 339 (2009), b2981.

10 International Committee of Medical Journal Editors, "Clinical Trial Registration," published simultaneously in the member journals, among them *Canadian Medical Association Journal* 171 (2004), pp. 606 f.: "This policy applies to any clinical trial starting enrollment after July 1, 2005."

11 World Health Organization, "New Study Estimates 151 000 Violent Iraqi Deaths since 2003 Invasion," press release, January 9, 2008 (Iraq Family Health Survey, 9,345 households; 104,000 to 223,000;

None of the three numbers is invented. Each method, in its strongest form, has good reasons. The count from reports contains only deaths that demonstrably occurred, with place and date; it underestimates, because many deaths appear in no report, especially where no reporters go. The surveys also reach the dead whom no one reported on; they infer from a few thousand households to a whole country, and their result depends on which areas got into the sample, how safely the interviewers could work there, and how precisely those interviewed remember. The smaller the sample, the wider the range.

Whoever takes only one number from this, the one that suits his own side, has not weighted but chosen. A calibrated judgment is one that knows the methods and takes the range seriously: that the count from reports is a lower bound, that the surveys lie far above it and differ from one another by a multiple, and that the question of the exact number therefore remains open, but the question of the order of magnitude not entirely. For present-day wars the same holds in sharper form. The numbers of the warring parties are determined by interests, those of the independent counters by their methods; together they say more than each by itself.

5.4 The Question for Every Number

What has been checked in this part about numbers can be put into four questions that belong together, because each starts at a different point between the world and the number. The first asks about the base rate: how frequent is that of which the number shows a case? The second about the denominator: of what is a part meant, and how large are the magnitudes behind a ratio? The third about the threshold: what counts as a case at all, and does the picture change when one moves the boundary? The fourth about selection: who could be counted, what was published, and is there a list that was fixed before the result? None of these questions requires mathematics. They require taking a number not as an answer but as the result of a process one can look at.

about three times as many as the count by Iraq Body Count from media reports, about four times fewer than the smaller survey of 2006), published as Iraq Family Health Survey Study Group, "Violence-Related Mortality in Iraq from 2002 to 2006," *New England Journal of Medicine* 358 (2008), pp. 484–493. The smaller survey: Gilbert Burnham, Riyadh Lafta, Shannon Doocy, Les Roberts, "Mortality after the 2003 Invasion of Iraq: A Cross-Sectional Cluster Sample Survey," *The Lancet* 368 (2006), pp. 1421–1428 (1,849 households; 601,027 violent deaths up to July 2006, range 426,369 to 793,663).

Numbers have the advantage that one can put these questions to them. Pictures, stories, and impressions have no denominator one could look up, and yet most of what people know about foreign countries arises from them.

Further reading: *Organization* (2026), Ch. 6 (what measuring does in institutions), *Truth* (2025), Ch. 19 (the feedback effect of measuring)

Summary

What gets counted has often already been selected before anyone calculates. In 1936 the Literary Digest polled ten million people from telephone directories and car lists, got 2.4 million answers, and predicted the wrong winner; quantity does not make a selection error smaller, only more precise. Of 74 studies on antidepressants submitted to the American regulator, half were positive by its assessment, but in the scientific journals 94 percent; this could be discovered because a register had been fixed before the result. The dead of the Iraq War were counted from reports and estimated in two surveys that differed from one another by a multiple; each method has good reasons, and a calibrated judgment is one that knows the methods and takes the range seriously. Four questions can be put to every number: about the base rate, the denominator, the threshold, and the selection.

PART III: FOREIGN COUNTRIES

Chapter 6: What One Knows from Afar

6.1 Every Channel Has Its Filter

Of Brazil, many people who have never been there know the carnival, the poor neighborhoods on the hillsides of Rio, and football. Of India, the films from Bombay, the poverty, and the programmers; of Japan, the overcrowded trains, the cherry blossom, and the robots; of Germany, the Oktoberfest, the cars, and the history of the twentieth century. None of this is invented. But each of these pictures has traveled a path on which it was selected – a path that has little to do with the frequency of what it shows.

A film needs a story, and a story needs a conflict, an exception, something that does not happen every day. It is set in places one recognizes, in the capital, in front of the well-known buildings; the country outside appears as landscape or not at all. A news report reports what departs from the usual: the attack, the election, the flood, the record. What circulates on social networks is what gets shared, and what gets shared is what outrages, moves, or makes people laugh. Travel reports show what is worth seeing. Each of these channels filters differently, but they filter in the same direction, toward the unusual and toward a few places. That is why their distortions do not cancel out when one takes them all together. They add up.

Whoever gets the same picture from five channels feels confirmed, with reason, insofar as the picture contains facts. For the weight he has no reason. Five filters that ask about the same thing are, for the question of how frequent something is, not five witnesses but one.

6.2 Accounts Are Filtered

What acquaintances who have visited a country tell is closer to reality than a film, but it too is selected, namely by what took the teller there. Whoever travels on business tells of business, of negotiations, partners, and hotels; he has seen

offices, airports, and restaurants where one takes guests, and about all of this he can report precisely. How a family lives there, what a farmer lives on, or how long one waits for a doctor's appointment he does not know, and mostly he does not tell of it either, because no one asks him.

It is similar with other paths. A correspondent often lives in the capital, among other foreigners and educated people, and sees the country from there. Whoever has emigrated and writes home tends to write what he has succeeded in rather than what failed, and whoever has failed perhaps does not write at all. A tourist sees what has been set up for tourists. Each of these reports is honestly meant. But each is a sample drawn according to the purpose of the journey. Whoever hears it should ask: What did he do there, where did he stay, whom did he talk to? Then he knows what the report is evidence for and what not.

6.3 What Cannot Be Shown

Some things reach the distant observer on no channel. A smell cannot be broadcast, nor a taste, nor the heat at midday, the loudness of a street, or the humidity of the air. Pictures show them as a hint, and the observer fills them in from his own life, that is, wrongly. Just as little does the ordinary show itself, because there is nothing to report: how a working day goes, how long one travels to work, what one eats for lunch, what one gets annoyed about, how one deals with the authorities. Yet precisely this ordinary is what makes up the greater part of life there. What is missing in the pictures is thus not an edge but the middle.

From this follows a simple question one can put to every picture of a distant country: What would be boring there? What is so self-evident that no one would film, report, or tell it? From afar one mostly does not know the answer, but the question alone already shifts the weight. It reminds one that what one knows is what is worth telling, and that what is worth telling makes up only a small part of a life.

6.4 What Once Held

A picture, once formed, remains even when the country changes. Whoever formed his picture of the world's poverty around 1990, when according to the World Bank's figures more than a third of people lived in extreme poverty, and has not renewed it, still sees today a world in which that is so; in 2019 it was less than a tenth. The same holds for life expectancy, the number of children in a family, the spread of telephones or schools in a country one once got to know.

Whoever was there at a certain time or read about it knows the country at that time. For every picture one must therefore ask when it arose and what might have changed since. The memory of one's own journey has a date too.

6.5 Calibrating from Afar

Even without a journey, the picture of a country can be readjusted a good deal. The most important means is to lay side by side channels that filter by different things: beside the pictures, the numbers that measure the ordinary, that is, how people live there, what they live on, how old they get, and what they spend on their living; beside the reports from outside, the newspapers of the country itself, which have filters of their own, but different ones; beside the capital, the provinces. One can ask who in the country writes about the country, and whether there are voices of those who do not live in the capital and do not belong to the educated. And one can check one's own pictures for the time they come from.

What does not succeed from afar is the direct experience of the context in which all this stands: how the numbers show themselves in a day, a street, a face. For that one has to go there. But what one sees there is also a frame, and one that considers itself more complete than any other.

Further reading: *Truth* (2025), Ch. 19 (what one knows about foreign countries)

Summary

What one knows about a country from afar comes from channels that each filter: films need conflicts and well-known places, news needs the unusual, social networks what gets shared, travel reports what is worth seeing. Because they filter in the same direction, their distortions add up; for the question of how frequent something is, five filters pointing the same way are one witness. Accounts by acquaintances, too, are samples drawn according to the purpose of their journey. What cannot be shown at all are smells, tastes, heat, and the ordinary, and the ordinary is the greater part of life there; the question of what would be boring there shifts the weight back. Every picture also has a date. One calibrates from afar by laying channels with different filters side by side: numbers about the ordinary, the country's own press, the provinces beside the capital.

Chapter 7: On the Spot and the Regions

7.1 What Comes Across Only on the Spot

Whoever comes for the first time to a country he knew only from pictures is often surprised by things no one told him about. Not by the sights – those he knew – but by small things that are so self-evident to the locals that no one mentions them: steam rising from the streets, traffic that sounds different, plants and birds he does not know, smells for which he has no words. These surprises are a measure of how filtered the pictures were with which he came. What surprises him was missing for him, without his having missed it.

On the spot one also takes in what shows only in the whole: how fast people walk, how close they come to one another, how loudly they speak, how they look at one another, where one shops and what lies in the displays, how far the distances are. Above all one sees how it all hangs together. The film shows one neighborhood; on the spot one sees how it borders on the next, who goes back and forth between the two, and who does not. The numbers from the previous chapter take on a shape here. Whoever knows that in a country most people shop in small stores knows it differently once he has shopped there for a month.

And yet one's own observation is filtered too, and its filter is harder to notice than that of a camera, because one does not have it before one but is it oneself. One lives in a particular neighborhood, one follows warnings about which area to avoid, one rides through a neighborhood on the bus and enters no house. One meets mainly people who deal with foreigners, at the hotel, in the store, at work. One is there in one season, and on holidays a city looks different from on working days. And the first impression, like the number from the wheel of fortune, weighs more than it deserves because it was the first. One's own observation is precious precisely because it shows what is missing in the pictures. But it is itself a sample and must be weighted like everything else.

7.2 Country, Region, City, Neighborhood

Whoever has been in a big city for two weeks knows a part of this city and not yet the country. What he saw there he must place: what of it belongs to the country, what to the region, what to the city, what to the neighborhood? The question is harder than it sounds, because one cannot read the levels off the single impression. That people walk fast in a city may be due to the country, to the big city, to the time of day, or to the fact that one was in the business district.

On closer look almost every country falls apart into regions that differ clearly in language, cuisine, faith, history, and manners, and these again into smaller ones. The journalist Colin Woodard has described eleven regional cultures for North America, which he traces back to the history of settlement and whose borders run across those of the states and even of the countries; some states have a share in several.[1] One need not follow his division in every line to see the finding: whoever describes "the Americans" mostly describes one or two of these cultures and takes them for the whole. The Indian constitution recognizes twenty-two languages, and whoever comes back from Kerala brings back a different India from someone who was in Punjab.[2] In Italy the north and the south, in Nigeria the north and the south, in China the coastal cities and the provinces of the west, in Portugal Lisbon and the uplands of the northeast are each like different countries for someone who knows both. Bavaria falls apart into Franconia, Swabia, and Old Bavaria, and in each of these parts there are again differences that the locals hear immediately.

From this follows no ban on speaking about a country as a whole. Some things hold throughout the country, because it has the same laws, the same schools, the same currency, and the same capital. But it follows that with every impression one must state the level at which one generalizes it, and that the level is not self-evident. What one says about "the Indians" an Indian from another region should not have to read as a description of a neighbor.

7.3 Calibrating a Stay

One can recalibrate one's own stay, just like the pictures from afar. One compares it with the impressions of others who were in another place at the same time or in the same place at another time. One asks locals what in one's own impression is typical of their country and what is not; they know the base rates of their country that a visitor lacks, but they too know mainly their own area and their own social stratum. One looks up what the statistics say about differences between the regions. And one distinguishes for oneself what one has seen often and what once, what one has seen oneself and what one has only heard.

---

1 Colin Woodard, *American Nations. A History of the Eleven Rival Regional Cultures of North America*, New York 2011; the eleven cultures (Yankeedom, New Netherland, The Midlands, Tidewater, Greater Appalachia, Deep South, El Norte, The Left Coast, The Far West, New France, First Nation) also in Woodard, "Up in Arms," *Tufts Magazine*, Fall 2013.

2 Constitution of the Republic of India, Eighth Schedule, with 22 languages in the version currently in force.

Nothing is thrown away in the process. What one has seen remains seen, even if it holds only for one neighborhood. Only its place in the picture changes: from a mark of a country to a mark of an area, a time, or a situation. That is more laborious than generalizing the impression or discarding it, but it is the only way in which it becomes knowledge.

A last turn reverses the chapter. Whoever comes back is asked what it was like, and tells. He tells what surprised him, what was funny, what was frightening, and not what was ordinary; he now becomes himself one of the filtered channels from the previous chapter. Whoever knows this can partly offset it by saying in addition what he did not see and what his report does not stand for.

Up to this point the weighting happened of itself, through cameras, channels, memory, and travel routes, without anyone wanting it so. The same means can also be used on purpose.

Summary

On the spot, what surprises is what no one announced because it is self-evident to the locals, and this surprise measures how filtered the pictures were. There one also sees how everything hangs together. But one's own observation is itself a sample: drawn by neighborhood, warnings, conversation partners, season, and first impression. What one sees belongs to the country, the region, the city, or the neighborhood, and the level cannot be read off the impression. Almost every country falls apart into regions that differ like countries: Woodard counts eleven regional cultures in North America across state borders, India's constitution recognizes twenty-two languages. One calibrates a stay by comparing it with other places and times, asking locals what is typical, and separating what was seen often from what was seen once; nothing is thrown away, only the place in the picture changes. Whoever comes home and tells becomes a filtered channel himself.

PART IV: WEIGHT AS A MEANS

Chapter 8: Propaganda

8.1 The Lie and Its Price

In the late summer and autumn of 1914, German troops advancing through Belgium and northern France killed some 6,500 civilians, many of them in shootings in towns and villages whose inhabitants they took for irregular fighters. Allied propaganda made more of it. Alongside the real crimes it spread stories of soldiers who had supposedly impaled infants and mutilated women, and much of this could never be substantiated or was invented. After the war this came to light, and the lesson many drew from it was: atrocity reports are propaganda. When, in the Second World War, reports arrived of the systematic murder of the Jews of Europe, many therefore took them to be exaggerated. They turned out to be essentially accurate.[1] The real crimes of 1914, too, were long denied after the war, until historians documented them from the records.

The lie in propaganda thus has a price that is paid twice. First it shifts the weight upward: the real crime, with invented features, becomes so heavy that it becomes proof of the wickedness of a whole people. Once exposed, it shifts the weight downward, and beyond the lie. Then every report of this kind weighs less, even the true one, even the next one, even the one about a crime greater than anything the lie invented. Whoever lies to give a report more weight damages the scale with which weighing will be done in the future.

A more recent case of this kind, the testimony about infants that Iraqi soldiers had supposedly taken from incubators in Kuwait in 1990, is treated in the book on truth. Such cases exist in many wars and on all sides. But they are not the rule, and precisely because they can be refuted, they are dangerous for whoever spreads them.

1 John Horne, Alan Kramer, *German Atrocities, 1914. A History of Denial*, New Haven 2001 ("some 6500 Belgian and French civilians," publisher's description, Yale University Press); United States Holocaust Memorial Museum, *Holocaust Encyclopedia*, article "Combating Holocaust Denial: Origins of Holocaust Denial": many British reports of German atrocities in Belgium turned out after the war to have been invented; "As a result of that experience, many people were skeptical of reports of mass murder operations during World War II."

Further reading: *Truth* (2025), Ch. 19 (the incubator testimony)

8.2 The Rule: Selection

The rule of propaganda is more modest and therefore more effective. Of the opponent it shows the bad, of its own side the good, and each single report can be true. The victims of one's own side have names, faces, and families; those of the other side are numbers, if they appear at all. One's own mistakes are isolated cases; the opponent's are signs of his nature. The opponent's violence appears without a prior history, one's own as a response.

None of this needs an authority that orders it. It arises even where editorial offices work freely, because people empathize more easily with those close to them, because one's own victims are easier to reach than those of the other side, and because an audience prefers to read what confirms its view. Where it is ordered, it is only more thorough. For the reader this makes little difference: in both cases he gets a picture made of true parts, whose weights are distributed by allegiance and not by what happened.

Experiment 3: Names and Numbers

Take a conflict that is currently being reported on, and a week of coverage from a source you usually read. Count how often victims of each side appear with a name, a picture, or a story, and how often only as a number. Then count how often the violence of each side is told with a prior history, as a response to something, and how often without.
If the distribution is strikingly unequal, you do not yet know who is right; perhaps one side really has more victims or more reasons. But you do know with what weight the parts reach you, and you can repeat the same count with a source from the other side. If it comes out as a mirror image, you have two filtered pictures and an idea of how large the filter is.

8.3 Balance as Staging

The subtler form shows both sides. A report lets supporters and opponents have their say, and no one can claim that a side was suppressed. The weighting then lies in the staging: the desired side gets the convincing speaker, the stronger image, and the last word; the undesired side gets the weakest representative

that could be found. The appearance of balance makes the weighting invisible, for whoever counts finds both sides represented.

The counter-form is just as much a distortion. It arises precisely from the wish not to stage. Whoever presents a voice for and a voice against on every matter, at equal length, gives a position that hardly anyone in a field holds the same weight as one that almost everyone holds. A report for the BBC's supervisory body concluded in 2011 that an overly rigid application of the rules on impartiality had given fringe opinions too much attention in science reporting, for instance on the safety of the measles-mumps-rubella vaccine and on genetically modified crops.[2]

Between the two errors there is no middle that one could calculate, but a duty to disclose the weighting. A report can say how many experts hold which position, what they rely on, and where the minority has a point that the majority has not yet answered. Then the reader can weigh for himself and check whether the stated weight is right. The weighting stays hidden when it lies in the choice of speakers or in the mere equal distribution of airtime. How large a majority is and how well founded, however, is itself a question of weighting, and the bodies that establish it have their own filters.

8.4 What One Thinks About

An American political scientist described the effect of the press in 1963 thus: it "may not be successful much of the time in telling people what to think, but it is stunningly successful in telling its readers what to think about."[3] That is a statement about weights. Which topic stays on the front pages for a week, which is mentioned briefly once, and which never, determines what a public considers urgent, long before anyone argues for or against.

Here too it helps to check the weights against something that does not come from the same stream. What cost the most people in my country their lives this year, what cost them the most money, what changed their everyday life the most, and how much about it was there to read in what I read? The answer may show

2 Steve Jones, report on the impartiality and accuracy of the BBC's science coverage, commissioned by the BBC Trust, 2011; reported in *Chemistry World*, "Report concludes BBC science is good but has weaknesses" (2011): an "over-rigid" application of the impartiality rules "gives 'undue attention to marginal opinion'," for instance on the safety of the MMR vaccine and on genetically modified crops.

3 Bernard C. Cohen, *The Press and Foreign Policy*, Princeton 1963, p. 13: the press "may not be successful much of the time in telling people what to think, but it is stunningly successful in telling its readers what to think about."

that attention was well distributed. It often shows that the ordinary and slow, which affects many, had less room than the sudden, which affects few. The reason need not be intent, for the sudden is new. The effect on the measure with which one separates the important from the unimportant is the same.

Staging and the choice of topics work with the distribution of what is shown. Another form works with its quantity.

Summary

Propaganda seldom lies, and where it lies, it pays twice: in 1914 German troops killed some 6,500 civilians in Belgium and France, Allied propaganda added invented atrocities, and when that came out, many took the accurate reports of mass murder in the Second World War for propaganda. Whoever lies to create weight damages the scale for future cases. The rule is selection: one's own victims with names, the others as a number, one's own violence as a response, the other side's without a prior history; this needs no order. The subtler form shows both sides and weights through staging. Its mirror image is false balance, which gives a fringe opinion as much room as what almost all experts hold. Between the two, the only help is to disclose the weighting. And the very choice of topics fixes what a public thinks about, long before anyone argues.

Chapter 9: Flooding

9.1 Quantity Instead of Persuasion

In 2016 a study by the RAND Corporation, a research institute that works mainly for the American armed forces, described Russian state propaganda with the image of a firehose full of falsehoods. Its marks, it said, were high volume and many channels, speed, continuity, and repetition, and a lack of commitment to reality as well as to consistency: what was claimed yesterday may be replaced today by its opposite.[4] Two years later an American campaign strategist described

4 Christopher Paul, Miriam Matthews, *The Russian "Firehose of Falsehood" Propaganda Model. Why It Might Work and Options to Counter It*, RAND Corporation, Perspective PE-198, Santa Monica 2016: "high-volume and multichannel," "rapid, continuous, and repetitive," "lacks commitment to objective reality," "lacks commitment to consistency."

his way of dealing with the press in his own country in a sentence that cannot be put politely: the way to deal with them is to "flood the zone with shit."[5]

The two cases come from different countries and political camps, and that is why they belong side by side here. Flooding is not a peculiarity of one regime or one direction. It is a technique open to anyone who can put more claims into circulation than others can check, and its aim is not to persuade. Its aim is the measure by which one is persuaded.

9.2 What Quantity Does to the Measure

Quantity works in three ways that reinforce one another. The first is repetition. What one encounters often becomes familiar, and the familiar feels frequent and credible, even when one knows where it comes from. The sense of frequency that the second chapter dealt with can thus be fed directly: not with arguments but with repetitions.

The second is leveling. Where twenty mutually contradictory claims circulate on every question, the well-documented one loses its weight, not because it has been refuted but because it is one among twenty. The measure that distinguishes between established, disputed, and unknown becomes flat. In the end everything seems roughly equally uncertain, and whoever has reached that point no longer believes the one who has the better reasons, but the one he trusts anyway, or no one.

The third is exhaustion. Claims can be made faster than they can be checked, and whoever refutes each one individually has used up his strength before the next wave comes. In addition, every refutation names the claim once more and thereby gives it attention. The study on the firehose itself drew the conclusion from this: "don't expect to counter the firehose of falsehood with the squirt gun of truth." It recommended warning people in advance and showing them how the technique works, instead of fighting each individual falsehood afterward; one should put raincoats on those at whom the hose is aimed.[6]

With this the question shifts. Against a single lie, checking the single statement helps. Against flooding it helps little, because the quantity aims precisely at overloading it. Only what keeps the measure itself stable can help.

---

5 Steve Bannon in conversation with Michael Lewis, reported in Michael Lewis, "Has Anyone Seen the President?," *Bloomberg View*, February 9, 2018: "the way to deal with them is to flood the zone with shit."

6 Paul, Matthews, *The Russian "Firehose of Falsehood" Propaganda Model* (2016): "don't expect to counter the firehose of falsehood with the squirt gun of truth"; recommendations to "forewarn" the audience and to put "raincoats" on those at whom the hose is aimed.

9.3 What Keeps the Measure Stable

The first means is slowness. The first reports about a major event are often incomplete and sometimes false, even without any intent, because no one yet knows what has happened. Whoever waits a day before forming a judgment loses little and gains much: what still stands after a day has been checked more than what circulated in the first hour. What still stands after a week, all the more. Flooding lives on the moment. Whoever measures it against time sees how much of it has vanished after a short while without ever having been retracted.

A track record also gives a hold. Whoever has a few sources whose hit rate he knows, because he has noted how often they were wrong and whether they corrected their errors, can hold on to them in a flood. That is no loyalty to a camp. Such a list includes sources of different directions; each stays on it as long as its track record holds.

Just as important is to distinguish the states of a question. The flood wants to bring everything into the state of the disputed. What helps against this is to keep hold, for oneself, of what is established in a matter, what is really disputed, and what no one knows, and not to mix these three compartments because many people are talking loudly. That many claims circulate on a question is a finding about the circulation, not about the question.

Finally there remains the source behind the source. Where a document, a recording, a data set exists, a look at it is worth more than a look at ten reports about it. It costs time, and so one cannot do it for everything. For what matters, it is the fastest way out of the flood.

None of these means requires that someone turn off the flood for one. They require that the individual tend his measure. Whether that is enough when a state or an alliance wants to defend itself against flooding and deception from outside, or whether it must then set the weights for its citizens itself, has meanwhile become a question of security policy.

Further reading: *Truth* (2025), Ch. 18 (the unearned claim to be checked and flooding)

Summary

Flooding aims not at persuading but at the measure by which one is persuaded. In 2016 a RAND Corporation study described Russian propaganda as a firehose of quantity, repetition, and indifference to reality and consistency; in 2018 an American campaign strategist recommended flooding the zone in his own country. Quantity works through repetition, which makes the familiar seem frequent; through

leveling, which makes established, disputed, and unknown indistinguishable; and through the exhaustion of those who refute each claim individually. That is why what helps against it is less the checking of the single statement than what keeps the measure stable: slowness, a track record of a few sources of different directions, the separation of the established, the disputed, and the unknown, and a look at the source behind the source. The RAND study itself recommended forewarning the audience instead of fighting each falsehood afterward.

Chapter 10: Cognitive Warfare

10.1 The Security Argument

In April 2020, in the first months of the pandemic, the European Union's External Action Service reported that pro-Kremlin sources and Russian state media were running a coordinated campaign to undermine the Union's response to the crisis and to sow confusion about the origin and consequences of the virus; official Chinese bodies were trying, in an equally coordinated way, to deflect the blame for the outbreak from themselves.[7] Several states explicitly describe, in their doctrines, the public opinion of other countries as a field on which battles are fought. Whoever is responsible for the security of an open society cannot pass over this. Precisely because it is open, its citizens are easy to reach, and precisely because they vote freely, it pays for foreign powers to act on what they take to be true.

In 2021 the NATO member states declared that hostile information activities, including disinformation, belonged among the threats against which they had to become resilient, and that this resilience was a national responsibility.[8] A study produced on behalf of one of the alliance's commands, which by its own account does not represent the command's position, went further. It described the human mind as a field of conflict in its own right, encompassing the whole human environment, friend as well as foe, and every user of modern information

7 European External Action Service, *EEAS Special Report Update: Short Assessment of Narratives and Disinformation around the COVID-19/Coronavirus Pandemic (Updated 2–22 April)*, April 27, 2020: "pro-Kremlin sources and Russian state media continue running a coordinated campaign"; on China, the coordinated shifting of blame and "covert Chinese operations" on social networks.

8 *Strengthened Resilience Commitment*, NATO summit in Brussels, June 14, 2021, para. 4 ("Resilience is a national responsibility and a collective commitment") and para. 7 ("hostile information activities, including disinformation").

technology as a potential target.[9] The word for this is cognitive warfare. The strongest version of the thought behind it runs: if the perception of one's own population is attacked, one must defend it, and whoever wants to defend must understand the field on which the attack takes place. That is right. The question is how to defend.

Further reading: *Truth* (2025), Ch. 17 (when the state steers checking; the sources on cognitive warfare in Appendix D)

10.2 Two Ways

Roughly, two ways can be distinguished. The one wants to bring the picture the population forms into the right direction itself: strengthen desired representations, weaken undesired ones, set its own narrative against the foreign one, direct attention. The other wants to strengthen the population's capacity to weight for itself: to notice when a picture made of true parts is built askew, to distinguish quantity from justification, to ask about denominator, selection, and source. The first way steers perception; the second strengthens calibration.

Both ways are taken by people who want to protect their country, and neither presupposes bad intentions. The difference lies in how they are built, and from this it can be shown why the second is the better defense, also against foreign manipulation and also from the standpoint of security alone.

10.3 Why Steering Defends Less Well

Steering perception creates a main switch. Whoever fixes what the population is to consider important, frequent, and true fixes the weights for everyone, and if he is wrong, everyone is wrong with him. Security doctrine knows the problem under another name: a point whose failure paralyzes the whole system is the first thing an opponent looks for. If it is deceived, infiltrated, or merely badly advised, it spreads the error with the whole reach it was given for defense.

Steering, moreover, works with the same means it is directed against: with selection, emphasis, repetition, and the staging of balance. For the citizen, weights steered from outside and from inside cannot be told apart as long as he does not

9 François du Cluzel, *Cognitive Warfare*, NATO Innovation Hub, Norfolk, January 2021 (sponsored by Allied Command Transformation, by its own account not the position of the command or of the member states), p. 6: "Any user of modern information technologies is a potential target"; p. 29: "The Human Domain of Operations refers to the whole human environment, whether friend of foe" ("of" thus in the original).

know that steering is taking place. A population used to following the signals of an authority instead of weighting for itself is therefore also easier to reach for foreign steering. An opponent then need not persuade; he need only imitate the signals one is used to following.

Steering becomes known sooner or later, in open societies mostly sooner. Then what the eighth chapter showed with the atrocity reports of the First World War happens: trust falls not only for the steered message but for everything that comes from the same source, including what is true. But defense lives on trust. In an emergency a population must believe a warning, a call to seek shelter, a report about where there is danger and where not. An authority that steered perception in quiet times finds less belief at that moment than it needs, and an opponent who knows this need only recall the earlier steering.

Finally, steering acts back on those who steer. Whoever weights every day what others are to see runs the risk of weighting his own picture of the situation by the same measure. Institutions that dose outward see less well inward, not at once, but over time.

Further reading: *Truth*, Ch. 17 (true inward, false outward)

10.4 Why Calibration Defends Better

A population that can weight for itself has no main switch. Its calibration is distributed over many heads, and whoever wants to shift it must reach many individually, each of whom carries out his own tests. That does not make it invulnerable. But it makes it expensive to deceive, and therein lies a deterrent that no central defense can offer.

The study on the firehose, from the circle of the American armed forces, itself came to this conclusion. Against the flood, it said, refuting every falsehood helps less than forewarning the audience and showing it how the technique works, that is, putting raincoats on those at whom the hose is aimed. A raincoat does not steer. It protects whoever wears it, no matter who is holding the hose at the moment. Similar is the mandate of a Swedish authority set up in 2022 for psychological defense: it is to identify and analyze improper information influence directed against the country and help counter it, and in doing so to protect the open, democratic society and the free formation of opinion.[10] There the defense is

10 Förordning (2021:936) med instruktion för Myndigheten för psykologiskt försvar, adopted on October 15, 2021, in force since January 1, 2022, § 1 (strengthening "befolkningens motståndskraft"), § 2 no. 1 ("identifiera, analysera och kunna lämna stöd i bemötandet av otillbörlig informationspåverkan"), and § 2 last paragraph ("värna det öppna och demokratiska samhället och den fria åsiktsbildningen").

directed at the procedure by which foreign actors exert influence – concealed origin, forged senders, coordinated deception – not at the opinions within the country.

An index that compares the resilience of European societies against disinformation measures it by press freedom, by education, above all by reading ability, and by people's trust in one another; in 2026 Denmark, Finland, Ireland, and the Netherlands shared the top place.[11] The index itself says that it does not measure media literacy but conditions that predict it. None of these conditions is a form of steering. They are the conditions under which people can weight for themselves.

For the institutions that defend a country, much remains to be done, and none of it requires steering perception. They can expose who is behind covert campaigns and present their evidence in such a way that others can check it; clarifying the origin of a message is something other than deciding about its content. They can forewarn the population about which procedures are currently being used. They can publish their own pictures of the situation with denominator, range, and uncertainty, including their own errors, and thereby strengthen the measure against which foreign claims must be measured. And they can treat education, a free press, and independent counts as part of defense, for that is what they are.

10.5 The Limit in an Emergency

One objection remains: in an emergency there is no time to calibrate a population. Whoever must go to the shelter needs an instruction, not a weighing-up. That is true, and it does not speak against the second way. An instruction about what to do is something other than steering what one takes to be true and important. It is followed if the body that gives it has earned trust beforehand, and it earns this trust in quiet times by weighting as the matter requires and disclosing its weights. What must be said in an emergency can afterward be publicly checked and justified. The calibration of citizens is thus not an alternative to leadership in a crisis but its precondition.

11 Marin Lessenski, *Measuring Vulnerability of Societies to Disinformation: The Media Literacy Index 2026*, Open Society Institute – Sofia, January 2026: 41 European countries; weighting media freedom 40, education 45, trust 10, e-participation 5 percent; "an instrument for measuring if not media literacy itself, then predictors of media literacy"; Denmark, Finland, Ireland, and the Netherlands jointly in first place.

The capacity to weight, however, is called for not only among citizens and authorities but everywhere people decide together: in committees, in professional societies, and wherever a consensus is established.

Summary

Foreign disinformation is documented, for instance by the European Union's External Action Service during the pandemic, and open societies are exposed to it because they are open. The NATO states call resilience against hostile information activities a national responsibility; concept studies describe the human mind, friend as well as foe, as a field of conflict. It can be defended in two ways: by steering perception or by strengthening calibration. Steering creates a main switch, works with the opponent's means, accustoms the population to signals an opponent can imitate, uses up – as soon as it becomes known – the trust on which warnings in an emergency depend, and clouds the picture of the situation of those who steer. Calibration is distributed and therefore expensive to deceive; the RAND study recommended forewarning the audience, and the Swedish authority for psychological defense is to counter foreign influence and protect the free formation of opinion. Defenders can expose origins, forewarn, publish their own pictures of the situation with their uncertainty, and treat education and a free press as part of defense. Instructions in an emergency remain necessary; they are followed where trust was earned beforehand.

PART V: WEIGHTING TOGETHER

Chapter 11: Committees, Experts, Consensus

11.1 Eight Hundred Estimates

At a livestock show in Plymouth in the southwest of England in 1907, one could estimate the weight of an ox after it had been slaughtered and dressed. Around eight hundred visitors handed in a ticket, and 787 of them were legible. The naturalist Francis Galton analyzed them. The middle estimate, the one with as many above it as below it, was 1,207 pounds. The ox weighed 1,198.[1]

None of the visitors was an expert on the weight of slaughtered oxen, and most were far off. Together they were almost right, because their errors went in different directions and canceled one another out. This succeeds under one condition: each estimated for himself, on his ticket, without knowing what the others had written. Had one of them first called out loudly what he thought, and had the others followed him, the errors would no longer have been independent. They would not have canceled out but added up, like the filters in the sixth chapter.

Many committees work exactly the other way round. First the one who chairs or carries the most standing speaks, then the others, and their judgments gather around the first like the estimates around the number from the wheel of fortune. The committee then has not the knowledge of ten heads but that of one, confirmed by nine. A simple remedy follows from Galton's finding: each records his judgment in writing before the discussion begins, and the discussion starts with the deviations, not with the opinion of the highest-ranking member.

1 Francis Galton, "Vox Populi," *Nature* 75 (1907), no. 1949, pp. 450–451 (West of England Fat Stock and Poultry Exhibition, Plymouth; 787 usable cards; median 1,207 pounds, actual dressed weight 1,198 pounds).

11.2 The Knowledge in the Room

On the evening before the launch of the space shuttle Challenger in January 1986, engineers of the manufacturer of the solid rocket boosters warned that the boosters' sealing rings were not safe in the expected cold. Their company at first recommended not launching. When the space agency objected, the company's management deliberated again. One of those involved remembers how a senior manager asked the head of engineering to take off his engineering hat and put on his management hat. The head of engineering then voted against his engineers. The shuttle broke apart on January 28 shortly after launch, and seven people died.[2]

The knowledge that mattered was in the room. It was not missing; it was weighted, and the weight was decided not by the data but by the role: a judgment that would have weighed heavily as a technical judgment weighed less as one among several management considerations. The request to change hats is in this sense a precise image. It did not change the knowledge but the scale on which it was weighed.

The lesson from this concerns every institution in which experts prepare and others decide. The weight of a technical objection must not depend on who has the higher rank in the room or which role currently applies. It must depend on what the objection shows, and whoever overrides it must say why he considers it weaker than the reasons against it.

Further reading: *Organization* (2026), Ch. 5 (what those who prepare and those who decide owe each other), *Ethics – The Good We Already Do* (2026), Ch. 15.8 (what one must know)

11.3 The Appointed Dissent

In 1587 Pope Sixtus V established an office whose holder, in proceedings for beatification and canonization, was to put forward everything that spoke against the candidate: weaknesses in his life, doubts about the miracles reported, exaggerations by his advocates. Popular speech called him the devil's advocate. In 1983 the procedure was reformed, and the office then lost influence.[3]

2 National Public Radio, "40 Years after Challenger, Lingering Guilt and Lessons Learned," January 2026, from the recollection of the engineer Brian Russell: "Bob, it's time to take off your engineering hat and put on your management hat." Before this, the company had recommended not launching and changed course after NASA objected.

3 *Encyclopaedia Britannica*, article "devil's advocate": establishment of the office (promotor fidei) by Sixtus V in 1587; task of presenting unfavorable facts; reorganization by John Paul II in 1983

The thought behind it reaches beyond the Church. Where all involved wish a matter to succeed, what speaks against it is easily weighed too lightly. An appointed counterweight gives it a place where it cannot be passed over without this being noticed. A committee can commission someone to present the strongest version of the opposing position, or ask a body that gains nothing from the outcome for a second view.

The means has a danger that it creates itself. If dissent becomes a role, everyone gets used to hearing it as a routine exercise, and so it loses precisely the weight it was meant to create. An appointed dissent carries weight only as long as it can actually win, that is, as long as a decision occasionally turns out differently because of it, and as long as this can be looked up.

11.4 Consensus About What

In 2021 the Intergovernmental Panel on Climate Change wrote in the first sentence of its headline statements that "it is unequivocal that human influence has warmed the atmosphere, ocean and land." For the question of how strongly the earth warms in the long term if the carbon dioxide content of the air doubles, it gave three degrees as the best estimate, with a range narrower than in the previous report.[4] Between these two statements lies a difference in certainty, and the report expresses it: the one counts as unequivocal, the other as an estimate with a range.

A consensus is therefore not a single weight but a series of weights for different questions. That the earth is warming and that humans have a part in it is something other than the question of how strongly it will warm further, and that again is something other than the questions of what consequences this has in which place, what a measure costs and what it brings, and which costs one wants to impose on whom. For the first question there is an agreement that is hard to overestimate; for the middle questions there are ranges that the experts themselves state; the last are not questions for experts alone but also questions of weighing, in which experts have no greater voice than other citizens.

Whoever appeals to a consensus therefore owes a statement of what it is about. Whoever disputes it owes the same statement; otherwise he disputes

with the constitution *Divinus perfectionis Magister*, after which the promotor fidei had "little sway over the proceedings."

4 IPCC, *Climate Change 2021: The Physical Science Basis*, contribution of Working Group I to the Sixth Assessment Report, headline statements of the Summary for Policymakers, A.1: "It is unequivocal that human influence has warmed the atmosphere, ocean and land"; A.4: "a best estimate of equilibrium climate sensitivity of 3°C, with a narrower range compared to AR5."

all levels at once, including those he does not mean at all. The same holds for every field in which committees establish agreement, in medicine, in economics, in history. The more precise the question, the more precisely one can say how much a consensus weighs, and the less it can be used as a lever for something it does not cover.

Committees are slow, experts are expensive, and both have their own filters. For a short while now there has been a conversation partner that answers everything quickly and cheaply and weights in doing so, without one seeing it.

Summary

Judging together is better than judging alone if the judgments are independent: in 1907, 787 visitors to a livestock show estimated the weight of an ox at a median of 1,207 pounds; it weighed 1,198. Whoever first hears the highest-ranking member turns ten judgments into one; that is why each records his judgment before the discussion begins. Before the Challenger launch in 1986 the engineers' warning was in the room and was given too little weight when a manager asked the head of engineering to exchange his engineering hat for his management hat; the weight of an objection must not depend on rank or role. An appointed dissent, like the office of the promotor fidei from 1587 on, gives the counterweight a place, but it carries weight only as long as it occasionally wins. A consensus is a series of weights for different questions: the Intergovernmental Panel on Climate Change calls human influence on warming unequivocal and its extent an estimate with a range. Whoever appeals to a consensus or disputes it owes a statement of what it is about.

Chapter 12: Artificial Intelligence

12.1 What Was Written

A language model has learned what people have written. When it says what is usual, what most experts think, or how things go in a country, it reproduces how frequently something occurred in these texts. That is a different frequency from the one in the world, and it is filtered in the same way as the channels from the sixth chapter. More is written about the unusual than about the usual, more about the capitals than about the villages, more in the languages in which much has been published and digitized than in the others, and more by those who write

than by those who do not. What a model knows about a country in which little is written, it knows largely from what others have written about this country.

A model can therefore be a very well-read conversation partner and yet have the same skewed measure as a person who has read much and seen little. Unlike with the person, one cannot tell from its answer where it gets its weights. The answer comes in a calm, confident tone whether it rests on a thousand sources or on three, and this tone is not a weight but a property of the language the model has learned.

12.2 The Tendency to Agree

In addition there is a tendency that stems from the way such models are refined. People rate their answers; the model learns to give answers that are rated well. Studies have shown that widely used assistants of this kind tend to tell the questioner what he wants to hear, and that human raters often prefer such answers.[5] For calibration this means: whoever puts a question in such a way that his own presumption is contained in it easily gets it confirmed, with reasons that sound good. The model becomes a mirror that throws back the questioner's measure instead of checking it.

What helps against this tendency is what also helps with people, only more deliberately. One asks without revealing one's own presumption. One has the model present the strongest version of the opposing position, and then the strongest version of one's own, and compares. One puts the same question once from the one side and once from the other and looks to see whether the answer moves along with it. If it moves, one has learned little about the question and a good deal about the model.

12.3 One Channel for Many

There are many newspapers, and each has its own filter; their errors go in different directions and partly cancel out, like the estimates in Plymouth. Language models that are consulted by very many people are few. If millions readjust their measure of the usual and the disputed against the same few models, their judgments become more alike, and their errors run in the same direction. That is the

5 Mrinank Sharma et al., "Towards Understanding Sycophancy in Language Models," arXiv:2310.13548 (2023): five AI assistants consistently showed the behavior across four tasks; "sycophancy is a general behavior of RLHF models," probably also because human ratings favor agreeable answers.

situation of the committee in which everyone speaks after the highest-ranking member, and that of the main switch from the tenth chapter, only without anyone having set it up on purpose. Whoever changes the weightings of a widely used model, whether through the selection of the texts from which it learns or through the rules by which it is to answer, changes the measure of all who trust it.

From this follows no advice to avoid models, but advice to use them like any other source: as one voice among several whose track record one knows. Whoever decides something important asks more than one model, checks numbers and quotations at the original source, and treats the statement "most experts think" (DE: „die meisten Fachleute meinen") from a model like the same statement from the mouth of an acquaintance: as a pointer one can check, not as a count.

12.4 What It Is Good For

As a tool of the second check, the one about distribution, a model can do precisely what the individual finds hard. One can ask it what is missing in a representation, which groups, periods, or regions it passes over. One can have it convert a percentage into frequencies, ask about the denominator, about the base rate, about other thresholds, about the method by which a number was obtained. One can ask it to name the strongest criticism of a study one likes. None of these questions asks the model for a decision, but for material that one checks oneself.

The same capacity has a reverse side. In seconds a model can write a text that names both sides and yet favors one, or a hundred versions of the same claim that sound different. Staging balance and flooding thereby become cheap. The remedy is the same as in the two chapters that described them: ask about the weighting instead of the number of voices, look at the source behind the source, and do not take quantity for weight.

Up to this point the subject has been weights that others set: cameras, channels, editorial offices, committees, states, models. In the end, however, each of these weights depends on the measure with which the individual reads, and this measure can be checked oneself.

Further reading: *Truth* (2025), Ch. 19 (language models as the sediment of their texts), *Organization* (2026), Ch. 11 (language models in institutions)

Summary

A language model reproduces how frequently something occurred in the texts from which it learned, and this frequency is filtered like every channel: toward the unusual, toward the capitals, toward the languages in which much was published. Its confident tone is not a weight. Because people rate its answers, widely used models tend to agree with the questioner; what helps is to ask without revealing one's own presumption, to have both sides presented in their strongest version, and to check whether the answer moves with the question. Because a few models advise very many people, their judgments become more alike, and whoever changes the weights of a model changes the measure of all who trust it. One therefore uses it as one voice among several and checks numbers at the source. As a tool of the second check it is well suited: it can look for what is missing, denominators, base rates, and counter-criticism. That it makes staging and flooding cheap calls for the same remedies as before.

PART VI: CALIBRATING ONESELF

Chapter 13: One's Own Scale

13.1 When a Judgment Is Calibrated

Whoever hears in the weather report that the probability of rain tomorrow is thirty percent can neither confirm nor refute the sentence the next day; it rains or it does not. It can be checked only over many days. A forecaster is calibrated in a precise sense if, on the days for which he gives thirty percent, it rains on about three out of ten, and on the days with seventy percent on seven out of ten. Studies have shown that meteorologists can state such probabilities reliably.[1]

That they in particular can do it has reasons that can be transferred. They judge the same kind of question every day, they learn the next day how it turned out, and their hits are counted. Most judgments in life have none of this. One judges different things, learns the outcome late or never, and no one counts. A measure that is never checked against outcomes can stand askew for decades without its owner noticing.

13.2 A Test

Experiment 4: Nine out of Ten

For each of the following quantities, give a range within which, by your conviction, the true value lies with ninety percent probability. The range should be neither so narrow that you would have to trust yourself with little, nor so wide that it no longer says anything, but such that with ten such ranges you would expect nine hits. Do not look anything up.

1 Allan H. Murphy, Robert L. Winkler, "Reliability of Subjective Probability Forecasts of Precipitation and Temperature," *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 26 (1977), pp. 41–47: "These results indicate that weather forecasters can formulate such forecasts in a reliable manner."

The mean distance from the earth to the moon. The height of Mount Everest. The circumference of the earth at the equator. The number of bones in the skeleton of an adult human. The speed of light in a vacuum.
The values: about 384,400 kilometers; 8,849 meters, according to the joint survey by China and Nepal in 2020; about 40,075 kilometers; 206; about 299,792 kilometers per second.[a] Count how many values lie within your ranges.
With five quantities the result is a small sample, and one miss may be chance. But whoever has only two or three hits is not alone: in experiments of this kind, the ranges that people consider ninety percent certain are mostly too narrow.[b] One trusts one's measure more than it deserves. Whoever, on the other hand, has five hits out of five may have chosen ranges wider than he would defend with a clear head, and that too is a deviation, only in the other direction.

---

a Moon: NASA, Moon Fact Sheet (semimajor axis 0.3844×10^6 km); Everest: 8,848.86 m, announced jointly by China and Nepal on December 8, 2020 (among others *Al Jazeera*, December 8, 2020); equatorial circumference from the equatorial radius of 6,378.137 km (NASA, Earth Fact Sheet); bones: National Cancer Institute, SEER Training Modules, "Divisions of the Skeleton"; speed of light: 299,792.458 km/s, fixed since 1983 by the definition of the meter.

b Marc Alpert, Howard Raiffa, "A Progress Report on the Training of Probability Assessors," in: Daniel Kahneman, Paul Slovic, Amos Tversky (eds.), *Judgment under Uncertainty: Heuristics and Biases*, Cambridge 1982, pp. 294–305; in summary David R. Mandel et al., "Effect of Confidence Interval Construction on Judgment Accuracy," *Judgment and Decision Making* 15 (2020), pp. 783–797: "the confidence intervals people generate are often too narrow or overly precise"; according to Budescu and Du (2007), ninety-percent ranges were too narrow, seventy-percent ranges well calibrated.

The experiment does not check what you know but how well your certainty fits your knowledge. That is calibration in the most precise sense. It can be practiced like any skill.

13.3 Keeping a Record

What meteorologists have by profession, anyone can set up for himself for a part of his judgments. One writes down what one expects, with a probability attached: that a project will be finished by autumn, that an election will turn out a certain way, that a dispute can be settled, that a number will rise next year. After a while one looks back. A few dozen such notes are enough to see whether one tends to

set one's certainty too high or too low, and in which areas. It may turn out that one does well on technical questions and badly on political ones, or the reverse.

Such a record of expectations is the calibration of calibration: checking not only the judgment about the world but the judgment about one's own reliability. It has a side benefit that is perhaps more important than the main benefit. Whoever knows that he writes down his expectations formulates them more precisely and can no longer reinterpret them afterward in such a way that he was always right.

13.4 The Symmetry Test

The hardest distortion of one's own measure is the one that runs with one's own wishes. A study that fits one's own conviction easily weighs more than one that contradicts it, even if both are equally well done. A single case on the other side counts as an exception, one on one's own side as a sign. What gets checked again is above all what is disturbing.

What helps against this is a test that swaps the labels. One takes a report that outrages or pleases one and in thought replaces those involved with those of the other side: the government one rejects with the one one votes for, the foreign country with one's own, the party with its opponent. Then one asks whether one would weight the report the same way, whether one would believe as quickly, check as strictly, be as loudly outraged. Where the answer is no, one has not discovered who is right, but one has discovered where one's own measure depends on allegiance instead of on the matter. The same test applies to studies: Would I accept this method if it had yielded the opposite?

The test has a limit. Not all situations are symmetrical; sometimes one side really has done more wrong or has better reasons. The symmetry test does not require judging both equally, but applying the same standards to both. What then turns out differently is different because the matter is.

13.5 The Strongest Opposing Side

A measure of one's own conviction can be gained only if one knows what it holds its ground against. Whoever knows the opposing side only from its weakest representatives – the loudest, the least precise, those one's own side laughs at – takes his own position to be stronger than it is, because he has never weighed it against its real counterweight. To calibrate here means to seek out the strongest version of the opposing position, in its own words and from its best representatives, and

only then to see how heavily one's own reasons weigh. Often one keeps one's conviction. But one then knows how sure one may be, and one can say what one would measure it by if one were wrong.

John Stuart Mill put this into one sentence in 1859: "He who knows only his own side of the case, knows little of that."[2]

Knowing the strongest opposing side requires, of course, that it exists and that it can speak. That is no longer a matter for the individual but for the space in which many calibrate together.

Further reading: *Truth* (2025), Ch. 19 and 11 (the tilted filter; when one was last turned around)

Summary

A judgment is calibrated when the stated certainty fits the outcomes: on the days for which a meteorologist gives thirty percent rain, it rains on about three out of ten, and studies show that weather forecasters can do this reliably because they judge daily, learn the outcome, and are counted. Most judgments lack this. In experiments, ranges that people consider ninety percent certain are mostly too narrow; one trusts one's own measure more than it deserves. Whoever writes down his expectations with probabilities and later looks back checks not only his judgment about the world but the one about his own reliability. The symmetry test swaps the labels and asks whether one would weight the same report about the other side equally; it requires not equal judgments but equal standards. And only whoever knows the strongest version of the opposing side learns the weight of his own conviction.

Chapter 14: Calibrating Together

14.1 Why Several Voices Are Needed

The estimates in Plymouth hit the mark because many estimated independently of one another. The errors of the channels from which one knows a distant country added up because they filtered in the same direction. Both together yield a condition without which calibration does not succeed: one needs voices

2 John Stuart Mill, *On Liberty*, London 1859, ch. II ("Of the Liberty of Thought and Discussion"): "He who knows only his own side of the case, knows little of that."

that filter differently and do not copy from one another. A scale can be checked only against a weight that was not weighed on the scale itself.

From this it follows that the diversity of newspapers, schools, disciplines, and opinions in a society is more than a courtesy toward those who think differently. It is the instrument with which a society checks its weights. Where only one channel speaks, or many that all say the same because deviating voices fear disadvantages, no one can establish any longer whether the common measure is right, not even those who set it. Pluralism in this sense is not an attitude one may have or not, but a condition of knowledge about weights.

But pluralism is not arbitrariness. Diversity serves calibration as long as the voices refer to the same thing and can be checked against it, like the blind men against the same elephant. A voice that cannot say what would refute it and gives no reasons that others can check contributes nothing to the common scale, however loud it is.

Further reading: *Truth* (2025), Ch. 20 (oriented pluralism), *Truth*, Ch. 16 (checking together needs procedures)

14.2 Weighting Disputed Questions Oneself

In fierce disputes of the present, whether about the climate, about vaccinations, or about a war, there is usually a prevailing picture and a counter-picture. Both appear as a whole. Whoever follows the one takes over all its weights; whoever follows the other, likewise. Only whoever breaks the whole down into questions that can be checked one by one can weight for himself, for the sub-questions have different degrees of certainty, and the two pictures conceal this equally.

For the climate, the eleventh chapter has already demonstrated the breakdown: from the question whether humans are warming the earth, through the question of how strongly, to the question of what should be done and who bears the costs. With vaccinations one can break things down in the same way. How well does a vaccine protect against severe illness, and in which group? How frequent are which side effects, again in which group? How high is the risk of the disease itself for people in this group? And what follows from this for a recommendation, an obligation, or a free decision? The first three questions are questions about frequencies, and for them everything holds that the second part said about base rates, denominators, and selection. The fourth is a weighing-up in which the freedom of the individual and the situation of others also count. The strongest version of the one side usually rests on the benefit for those most at risk and on what a disease does that spreads unchecked. The strongest version

of the other rests on the fact that the benefit for people at low risk can be small, so that even rare side effects carry weight, and on the fact that an obligation must justify more than a recommendation. Both versions can be right depending on the group, and whoever sets them side by side sees that their dispute is to a large extent a dispute about which group one calculates in.

With wars there is the additional fact that both sides have an interest in their picture that goes beyond the truth. Here too breaking things down helps. What is documented by sources that belong to neither of the parties: pictures whose place and time can be checked, documents, independent counts? What do the counts of victims say, by what method, with what range, as the fifth chapter showed with the Iraq War? Which prior history does each side tell, and which does it leave out? And what is simply not yet known? Whoever answers these questions for a war seldom arrives at a picture that entirely pleases either of the two sides. That is not a defect of the procedure but often its result.

14.3 No False Middle

Weighting for oneself does not mean seeking the middle between two camps on every question. If the check shows that a sub-question has been answered almost unambiguously, then the calibrated answer is almost unambiguous, even if one camp disputes it. If it shows that a question is open, then the calibrated answer is "open," even if both camps proclaim certainty. False balance, which gives a fringe opinion as much weight as what is well documented, is as little calibrated as camp thinking, which does not check the fringe opinion at all.

Calibration therefore also includes being able to say "I don't know" (DE: „ich weiß es nicht") and to keep a question open instead of closing it with the weight of one's own camp. Whoever can do this will win arguments less often. But he will less often be tied to a side whose weights he has never checked.

14.4 Weighting with Others

The strongest help in calibrating are people who weight differently and can give reasons for it. A conversation with them is fruitful if both say beforehand what would change their minds: which number, which finding, which source. Then one no longer argues about the result but about how one could check it, and it often turns out that one agrees more about the facts than about their weight, or more about the weight than about the values by which one measures it. Both are progress, for only then does one know what one actually disagrees about.

Such a conversation presupposes that the other counts not as an enemy whose reasons one need only refute, but as an instrument that is gauged differently from one's own and therefore shows where one's own might deviate. This attitude is a protection against what this book itself could bring about. For whoever has learned how weights arise and are shifted can also use this knowledge to shift them.

Summary

Calibrating needs voices that filter differently and do not copy from one another; that is why the diversity of media, schools, and opinions is not a courtesy but the instrument with which a society checks its weights. It serves this purpose as long as the voices refer to the same thing and give reasons that can be checked. In disputes of the present, picture and counter-picture appear as a whole; whoever breaks them down into sub-questions with different degrees of certainty can weight for himself. For the climate, the breakdown runs from whether through how much to what to do; for vaccinations, from protection, side effects, and the risk of the disease in a particular group to weighing recommendation against obligation; for wars, from what is independently documented through counts and prior histories to what is not yet known. Calibrating does not mean seeking the middle; it includes saying "I don't know." (DE: „ich weiß es nicht") What helps most is conversation with people who weight differently, if both say beforehand what would change their minds.

Chapter 15: Inverted Calibration

15.1 Knowledge That Works in Both Directions

Whoever knows how a close-up enlarges a crowd, how a relative risk frightens, how a threshold reverses a picture, can recognize all this. He could also use it. A book about weights can therefore be read against the grain; no sentence can prevent that. But there is a reason why such knowledge is of more use to the one who wants to protect himself than to the one who wants to deceive. Whoever shifts weights needs an audience that does not know how it is done; for him the techniques have long been known. Whoever wants to recognize them mostly learns of them only once he has fallen for them. The more people know how distorted weights arise, the less they work. That was the advice of the study

on the firehose, to forewarn the audience, and it is the reason why this book describes the procedures, always from the side of the one who wants to notice and correct them.

More dangerous than the use of the techniques described is their inversion: that the means of calibration themselves, the questions of this book, are used in such a way that they no longer check but secure a result that was fixed beforehand. This seldom happens on purpose. Mostly it happens to people who are convinced that they weight better than others.

Further reading: *Truth* (2025), Ch. 17 (inverted principles), *Truth*, Ch. 18 (when the method itself is falsified)

15.2 The Questions of This Book, Inverted

The question “What is missing?” is the most important tool of this book and the easiest to invert. It is inverted when every absence counts as proof: this is not reported, so it is being concealed, so it is true and important. A question put this way can no longer be answered by anything, for even the answer one gets then counts as part of the cover-up. One recognizes the inversion by the fact that no one can state what one would have to see if nothing were hidden. The remedy is to demand this statement and then to go and look.

The base rate can be inverted by using it to dismiss every single case: this is rare, so it does not count. But a base rate distributes attention and the burden of proof, not truth; the single case is itself a finding, and if such cases accumulate, they change the rate. One recognizes the inversion by the fact that the rate is pointed to where the case would have to be checked.

Denominator and threshold can be chosen so that the desired picture emerges, and precisely with the claim to accuracy: I am only showing the numbers. One recognizes it by the fact that only one threshold, one period, one denominator appears, and the remedy is the question from the fourth chapter, whether the picture holds when one changes them. Whoever does not want to show the other versions has a reason for it.

The symmetry test can be inverted in two directions. In the one, equal standards become an equal judgment, and every matter gets its two equally strong sides, even where the evidence is unequal. In the other it becomes a defense: every criticism of one’s own side is followed by a reference to the offenses of the other, and with that the criticism is disposed of without having been checked. Both miss what the test requires, namely to apply the same standards to both and then to see what turns out differently.

The language of calibration itself can also be inverted. Whoever reproaches others with distortion, selection, neglected base rates and never checks himself for them has turned a tool into a weapon. And where the strengthening of judgment is announced and the drilling of particular judgments is pursued, calibration has become the steering that the tenth chapter dealt with, only under another name.

15.3 How One Recognizes the Inversion

The inverted forms have common signs, and they can be checked in every text, every training course, and every speech, including this book. The first is one-sidedness: the tool is applied in only one direction, against the others, never against one's own side. The second is irrefutability: there is no finding that could end the question or change the judgment. The third is breaking off: the check stops exactly where it would become dangerous to one's own result. The fourth is the direction in which the listener is brought: does he become more independent, can he weight better on his own afterward, or does he need the one who explains the weights to him more than before?

The four signs belong together. All indicate a result that was fixed before the check, and a check that now serves only to protect it. Where they appear together, the question is no longer who is right but whether any checking is still going on at all.

15.4 How to Protect Oneself and Others

The surest protection comes from applying it to oneself, first and most strictly. Whoever puts the questions of this book first to his own convictions, to the weights of his own side, to his own favorite numbers, can hardly use them as a weapon, for he has experienced how they cut.

Disclosure is just as effective. Distorted weights work as long as they are hidden; shown, they lose their force. Whoever gives numbers with denominator, range, and threshold, reports with what they do not show, and judgments with what would change them makes it hard for anyone to invert him, himself included.

Then there is spreading the knowledge. The more people know the procedures, the less they pay. Whoever helps others with this should show them how to weight, not how they should weight. Whoever hands his weights over to others as the right ones, even with the best intentions, makes himself into the main

switch this book warns against, and whoever passes on his questions distributes calibration over more heads.

Where many decide together, forms help that keep the weighting visible: proposals that name the possibilities rejected, registers set up before the result, committees in which each first judges for himself, counts that are independent of the parties. They do not prevent inversion, but they make it visible, and what is visible can be corrected.

What none of these means of protection provides is a certainty that one is weighting correctly oneself. Therein lies the last limit of calibration.

Summary

Knowledge about weights works in both directions, but it is of more use to the one who wants to protect himself than to the one who wants to deceive, for deception needs an audience that does not know the procedures; that is why this book describes them from the perspective of the one who wants to notice them. More dangerous is the inversion of calibration itself: the question "What is missing?" for which every absence is proof; the base rate used to dismiss every single case; denominators and thresholds shown only in the version that suits; symmetry as false equality or as a reference to the other side's offenses; the language of calibration as a weapon against others and as a name for steering. The inversion can be recognized by one-sidedness, irrefutability, breaking off the check where it would become uncomfortable, and by the listener becoming more dependent. Protection comes from applying it to oneself, disclosing the weights, spreading the questions rather than the answers, and forms that keep the weighting visible.

Chapter 16: Limits

16.1 Where No Feedback Comes

The meteorologists could check their certainty on thousands of days. Many of the most important judgments cannot be checked in this way, because the situation is rare or unique. Whether a country should start a war, slow an epidemic with harsh measures, or give up a currency is not decided a thousand times but once, and what would have happened had one decided differently one never learns. For such judgments there is no hit rate against which one's own certainty could be measured.

One can make do by looking for similar situations that occurred earlier, and for the rate at which they turned out one way or the other. That is the base rate from the third chapter, and with it comes its problem: the rarer the situation, the thinner and less certain the group of comparable cases, and the more depends on the choice of which cases one considers comparable. Where calibration becomes this thin, what follows from it is above all an attitude: ranges instead of points, decisions that can be reversed where that is possible, and the readiness to check early whether the assumptions hold.

16.2 What Is Missing and Does Not Show Itself

Some things are missing in such a way that no question finds them, because no one can put the question before something shows itself. Only in hindsight does it become visible that a presupposition that was self-evident to everyone was not correct. The general reservation that something might be missing helps little against this, for it applies equally to every judgment and does not say where. It is no reason to discard a particular judgment. It is a reason to keep the open places visible, not to set aside what is unexplained, and to look where something does not add up.

16.3 No Main Switch, No Conclusion

Calibration is never finished. The world changes, and with it the rates against which one's own measure would need checking; what was correct twenty years ago is perhaps a memory today. One's own measure also shifts without any change in the world, through what one is reading at the moment and whom one is listening to. What has once been readjusted must therefore be checked again, not constantly but regularly, and above all where much depends on it.

This duty to repeat can paralyze if one misunderstands it. To calibrate does not mean to check until no uncertainty remains; uncertainty always remains. It means to act with the best-checked measure one has, and to know how certain it is. Whoever waits until everything has been weighted has also weighted, only in favor of what happens without him.

A society in which many judge in this way has no main switch with which its measure could be reset. That makes it slower to reach agreement and harder to steer, even for those who mean well. But it also makes it hard to deceive. It can notice its errors because not everyone makes them at the same time.

16.4 This Book as a Selection

This book too is a selection. It must let itself be checked against what it demands of others. Many of its cases come from the United States and from Britain. That is not because more is distorted there, but because these cases are well studied and accessible in sources an author can check quickly: in English, freely online, often cited. That is exactly the filter the sixth and the twelfth chapters deal with. Whoever lives in another part of the world probably does not find the cases of his own country here. He should check whether the thoughts hold just as well for them. If they do not, the book is wrong at that point.

The same holds for the selection of topics. A book that mainly describes errors of weighting can create the impression that everything is wrongly weighted. Most of the weights with which people judge in everyday life are good enough for what they are needed for; otherwise no one would get through the day. One must calibrate above all where much depends on it, where the matter is distant, and where someone has an interest in the picture one forms.

16.5 The Step Back

In Firdos Square in Baghdad, what the close-up concealed showed itself when one stepped back: the farther one was from the pedestal, the more one saw. Therein lies what calibrating is at its core. One steps back from the picture that reaches one and asks how large the frame is, what lies outside it, how frequent what is shown is, who selected it, and how heavily it weighs beside the rest. One readjusts one's own measure against something that does not come from the same stream. And one keeps it open for the next step back.

Further reading: *Truth* (2025), Ch. 10 and 20 (the error that has not yet shown itself; the truth process)

Summary

Where a situation is rare or unique, no feedback comes against which one's own certainty could be checked; the base rate of similar cases helps but becomes thin, and what remains are ranges instead of points, reversible decisions, and early checking. Some of what is missing shows itself only in hindsight; the general reservation does not discard any particular judgment, but it demands keeping open places visible. Calibration is never finished, because the world and one's own measure shift. It does not mean checking until no uncertainty remains, but acting

with the best-checked measure and knowing how certain it is. A society in which many judge in this way has no main switch: it is slower to bring to agreement and harder to deceive. This book too is a selection, with many cases from English-language sources because they were easy to check; readers elsewhere should check the thoughts against their own cases. At its core, calibrating is the step back from the picture, as in Firdos Square, and the question of how large the frame was.